

# MATHER SETS FOR SEQUENCES OF MATRICES AND APPLICATIONS TO THE STUDY OF JOINT SPECTRAL RADII

IAN D. MORRIS

**ABSTRACT.** The joint spectral radius of a compact set of  $d \times d$  matrices is defined to be the maximum possible exponential growth rate of products of matrices drawn from that set. In this article we investigate the ergodic-theoretic structure of those sequences of matrices drawn from a given set whose products grow at the maximum possible rate. This leads to a notion of Mather set for matrix sequences which is analogous to the Mather set in Lagrangian dynamics. We prove a structure theorem establishing the general properties of these Mather sets and describing the extent to which they characterise matrix sequences of maximum growth. We give applications of this theorem to the study of joint spectral radii and to the stability theory of discrete linear inclusions.

These results rest on some general theorems on the structure of orbits of maximum growth for subadditive observations of dynamical systems, including an extension of the semi-uniform subadditive ergodic theorem of Schreiber, Sturman and Stark, and an extension of a noted lemma of Y. Peres. These theorems are presented in the appendix.

## 1. INTRODUCTION

Let  $A$  be a  $d \times d$  complex matrix, and let  $\|\cdot\|$  be a norm on  $\mathbb{C}^d$ . The *spectral radius* of  $A$ , denoted  $\rho(A)$ , is defined to be the maximum of the moduli of the eigenvalues of  $A$ , and satisfies the well-known formula

$$\rho(A) = \lim_{n \rightarrow \infty} \|A^n\|^{\frac{1}{n}} = \inf_{n \geq 1} \|A^n\|^{\frac{1}{n}}$$

due to I. Gelfand. By analogy with this formula, given a bounded set  $\mathsf{A}$  of complex  $d \times d$  matrices we define the *joint spectral radius* of  $\mathsf{A}$  to be the quantity

$$\varrho(\mathsf{A}) := \lim_{n \rightarrow \infty} \sup \left\{ \|A_{i_1} A_{i_2} \cdots A_{i_n}\|^{\frac{1}{n}} : A_i \in \mathsf{A} \right\} = \inf_{n \geq 1} \sup \left\{ \|A_{i_1} A_{i_2} \cdots A_{i_n}\|^{\frac{1}{n}} : A_i \in \mathsf{A} \right\}.$$

The existence of this limit and its identity with the above infimum is a straightforward consequence of Fekete's subadditivity lemma, which we include in the appendix as Lemma A.1. The joint spectral radius was introduced by G.-C. Rota and G. Strang in 1960 ([51], later reprinted in [52]), and in the last two decades has attracted substantial research attention. This has dealt with its applications, which include control theory [1, 16, 24], wavelet regularity [18], numerical solutions to differential equations [23], combinatorics [5], and coding theory [41]; with algorithms for its computation and estimation [6, 22, 35, 44, 49], and questions of its theoretical computability [58]; and with its intrinsic properties as a mathematical function [8, 26, 34, 61, 62].

If  $\mathbf{A}$  is a compact set of  $d \times d$  complex matrices, then it is not difficult to see that  $\varrho(\mathbf{A})$  admits the alternative formulation

$$(1) \quad \varrho(\mathbf{A}) = \sup_{(A_{i_j})_{j=1}^{\infty} \in \mathbf{A}^{\mathbb{N}}} \limsup_{n \rightarrow \infty} \|A_{i_n} \cdots A_{i_1}\|^{\frac{1}{n}}.$$

That is,  $\varrho(\mathbf{A})$  is the fastest possible exponential growth rate of the sequence of partial products of an infinite sequence of elements of  $\mathbf{A}$ . In this article we are concerned with the problem of understanding, for a given set of matrices, which sequences achieve this fastest possible exponential growth rate. In some form this problem goes back to early research by J. Lagarias and Y. Wang [37], and by L. Gurvits [24], in which it was asked whether this maximal growth rate is always attained by a periodic sequence. The general problem of understanding the structure of these fastest-growing sequences of partial products has arisen not only in attempts to answer Lagarias-Wang and Gurvits' question - such as in the articles [13, 25, 33] - but also in control theory [16] and when studying the approximability of the joint spectral radius by spectral radii of finite matrix products [44].

In this article we investigate these orbits of maximal growth from an ergodic-theoretic perspective. Namely, we study those shift-invariant measures on spaces of sequences of matrices for which the pointwise asymptotic growth rate of the sequence of partial products is maximised. We then relate this to some more concrete notions of orbits of maximal growth. Our first theorem, Theorem 2.1, gives some equivalent formulations of the definition of the joint spectral radius in terms of invariant measures. Motivated by earlier research in the fields of Lagrangian dynamics and ergodic optimisation, we go on to consider an object analogous to the *Mather set* in Lagrangian dynamics, which consists of the union of the supports of these growth-maximising invariant measures. Our main result, Theorem 2.3, is a structural theorem which describes the properties of these Mather sets and their relationship with orbits of maximal growth. To motivate this study we give an application of Theorem 2.3 to the stability theory of discrete linear inclusions introduced by L. Gurvits [24], and use Theorem 2.3 to prove some new results relating to the theory of joint spectral radii.

## 2. NOTATION AND STATEMENT OF MAIN THEOREMS

Before describing our results in more detail, let us establish some notation and conventions. Throughout this document the expression  $\log 0$  will be interpreted as being equal to  $-\infty$ , so that  $\log$  extends to a continuous function from  $\mathbb{R}_{\geq 0}$  to  $\mathbb{R} \cup \{-\infty\}$ . We shall use the symbol  $\mathbf{M}_d(\mathbb{C})$  to denote the set of all  $d \times d$  matrices with complex entries. In the remainder of the article the symbol  $\|\cdot\|$  will be used to denote both the Euclidean norm on the finite-dimensional space  $\mathbb{C}^d$ , and also the norm on  $\mathbf{M}_d(\mathbb{C})$  induced by the Euclidean norm. More general norms on  $\mathbb{C}^d$  or  $\mathbf{M}_d(\mathbb{C})$  shall usually be denoted by  $\|\cdot\|$ . We shall call a norm  $\|\cdot\|$  on  $\mathbf{M}_d(\mathbb{C})$  *submultiplicative* if  $\|AB\| \leq \|A\| \cdot \|B\|$  for all  $A, B \in \mathbf{M}_d(\mathbb{C})$ ; this property holds in particular for every norm on  $\mathbf{M}_d(\mathbb{C})$  which is induced by a norm on  $\mathbb{C}^d$ . We shall say that a norm  $\|\cdot\|$  on  $\mathbb{C}^d$  is an *extremal norm* for  $\mathbf{A}$  if the induced norm on  $\mathbf{M}_d(\mathbb{C})$  satisfies  $\|A\| \leq \varrho(\mathbf{A})$  for all  $A \in \mathbf{A}$ . A set of matrices  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  shall be called *relatively product bounded* if there exists a constant  $C > 1$  such that  $\|A_{i_1} \cdots A_{i_n}\| \leq C \varrho(\mathbf{A})^n$  for all  $n \geq 1$  and  $A_{i_1}, \dots, A_{i_n} \in \mathbf{A}$ . It is not difficult to

show that a set  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  is relatively product bounded if and only if it admits an extremal norm [29, 51].

In order to study the set of sequences of elements of  $\mathbf{A}$  whose partial products grow at a given rate, it would be most natural to work directly with the sequence space  $\mathbf{A}^{\mathbb{N}}$ . However, for various reasons this approach is inconvenient. For example, in proving several of our main results we will find it useful, when considering a sequence of matrices  $(A_{i,j})$ , to pass to a corresponding sequence of lower-dimensional matrices  $(B_{i,j})$  derived from the sequence  $(A_{i,j})$ , whilst retaining an awareness of the structure of the original sequence. This is problematic since in general the two kinds of sequence will not be related by a bijective correspondence. Furthermore, in other applications we shall be interested in how certain structures change when the set  $\mathbf{A}$  itself is perturbed. For these reasons, rather than working directly with  $\mathbf{A}^{\mathbb{N}}$  we shall instead choose to work on abstract spaces of *index sequences* which are detached from any a priori relationships with sets of matrices.

Given a compact set of matrices  $\mathbf{A}$ , we shall say that  $\mathcal{I}$  is an *index set* for  $\mathbf{A}$ , or simply that  $\mathbf{A}$  is indexed in  $\mathcal{I}$ , to mean that  $\mathcal{I}$  is a compact metric space equipped with a continuous surjection  $\mathcal{I} \rightarrow \mathbf{A}$  taking the symbol  $i \in \mathcal{I}$  to the matrix  $A_i \in \mathbf{A}$ . In general this map will not be assumed to be injective. Given an index set  $\mathcal{I}$ , we shall use the symbol  $\Sigma_{\mathcal{I}}$  to denote the metrisable topological space  $\mathcal{I}^{\mathbb{N}}$  of all one-sided infinite sequences in  $\mathcal{I}$ , which we equip with the infinite product topology inherited from  $\mathcal{I}$ . We define the *shift transformation*  $\sigma: \Sigma_{\mathcal{I}} \rightarrow \Sigma_{\mathcal{I}}$  by  $\sigma[(x_i)_{i=1}^{\infty}] := (x_{i+1})_{i=1}^{\infty}$ , which is a continuous surjection from  $\Sigma_{\mathcal{I}}$  to itself. When  $\mathcal{I}$  equals to  $\{1, \dots, \ell\}$  for some integer  $\ell$  we shall simply write  $\Sigma_{\ell}$  for the set  $\Sigma_{\mathcal{I}}$ . In general we shall leave the structure of  $\Sigma_{\mathcal{I}}$  as a metrisable space implicit, but in the special case of finite index sets we shall sometimes find it convenient to equip each  $\Sigma_{\ell}$  with an explicit metric. In these cases we shall consider the metric on each  $\Sigma_{\ell}$  given by

$$d[(x_i), (y_i)] := 2^{-\inf\{n \geq 1: x_n \neq y_n\}},$$

where  $2^{-\infty}$  is interpreted as 0.

When a particular shift space  $\Sigma_{\mathcal{I}}$  is understood, we shall use the symbol  $\mathcal{M}$  to denote the set of all Borel probability measures on  $\Sigma_{\mathcal{I}}$ , and the symbol  $\mathcal{M}_{\sigma}$  to denote the set of all such measures which are invariant under  $\sigma$ . We denote the set of all ergodic  $\sigma$ -invariant measures on  $\Sigma_{\mathcal{I}}$  by  $\mathcal{E}_{\sigma}$ . We equip  $\mathcal{M}$  with the weak-\* topology, which is the smallest topology on  $\mathcal{M}$  such that the map  $\mu \mapsto \int g d\mu$  is continuous for every continuous  $g: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R}$ . With respect to the weak-\* topology  $\mathcal{M}$  is compact and metrisable, and  $\mathcal{M}_{\sigma}$  is a closed subset of  $\mathcal{M}$  (see e.g. [59]).

We use the following notation to describe the partial products of sequences of matrices associated to elements of  $\Sigma_{\mathcal{I}}$ . If  $\mathbf{A}$  is a compact set of matrices indexed by  $\mathcal{I}$ , then for each  $x = (x_i)_{i=1}^{\infty} \in \Sigma_{\mathcal{I}}$  and each  $n \geq 1$  we define

$$\mathcal{L}_{\mathbf{A}}(x, n) := A_{x_n} A_{x_{n-1}} \cdots A_{x_1}.$$

The function  $\mathcal{L}_{\mathbf{A}}: \Sigma_{\mathcal{I}} \times \mathbb{N} \rightarrow \mathbf{M}_d(\mathbb{C})$  is continuous and satisfies the following *cocycle relation*: for all  $x \in \Sigma_{\mathcal{I}}$  and  $n, m \geq 1$ ,

$$\mathcal{L}_{\mathbf{A}}(x, n+m) = \mathcal{L}_{\mathbf{A}}(\sigma^n x, m) \mathcal{L}_{\mathbf{A}}(x, n).$$

Since by construction we have  $\{\mathcal{L}_{\mathbf{A}}(x, n): x \in \Sigma_{\mathcal{I}}\} = \{A_{i_n} \cdots A_{i_1}: A_i \in \mathbf{A}\}$  for each  $n \geq 1$ , the joint spectral radius  $\varrho(\mathbf{A})$  is described in terms of  $\mathcal{L}_{\mathbf{A}}$  by the expression

$$(2) \quad \varrho(\mathbf{A}) = \lim_{n \rightarrow \infty} \sup \left\{ \|\mathcal{L}_{\mathbf{A}}(x, n)\|^{\frac{1}{n}}: x \in \Sigma_{\mathcal{I}} \right\},$$

and in view of (1) we further have

$$(3) \quad \varrho(\mathbf{A}) = \sup_{x \in \Sigma_{\mathcal{I}}} \limsup_{n \rightarrow \infty} \|\mathcal{L}_{\mathbf{A}}(x, n)\|^{\frac{1}{n}}.$$

Given a set  $\mathbf{A}$  indexed in  $\mathcal{I}$  and an ergodic measure  $\mu$  on  $\Sigma_{\mathcal{I}}$ , the subadditive ergodic theorem (given in the appendix as Theorem A.2) implies that for  $\mu$ -a.e.  $x \in \Sigma_{\mathcal{I}}$  one has

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| = \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(z, n)\| d\mu(z).$$

In view of this and (3) it is natural to ask whether, given such a set  $\mathbf{A}$  and index set  $\mathcal{I}$ , there always exists an ergodic measure on  $\Sigma_{\mathcal{I}}$  such that  $\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu = \log \varrho(\mathbf{A})$ . The answer to this question is given by the following fundamental result, which provides us with a range of dynamical descriptions of the joint spectral radius  $\varrho(\mathbf{A})$ :

**Theorem 2.1.** *Let  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  be a compact set indexed in  $\mathcal{I}$ , and let  $\|\cdot\|$  be a submultiplicative norm on  $\mathbf{M}_d(\mathbb{C})$ . Then*

$$\begin{aligned} \log \varrho(\mathbf{A}) &= \sup_{\mu \in \mathcal{M}_{\sigma}} \inf_{n \geq 1} \frac{1}{n} \int_{\Sigma_{\mathcal{I}}} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) \\ &= \inf_{n \geq 1} \sup_{\mu \in \mathcal{M}_{\sigma}} \frac{1}{n} \int_{\Sigma_{\mathcal{I}}} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) \\ &= \sup_{x \in \Sigma_{\mathcal{I}}} \inf_{n \geq 1} \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\|. \end{aligned}$$

*In the first two expressions every supremum over  $\mathcal{M}_{\sigma}$  is attained by some ergodic measure, and in the last expression the supremum is attained for at least one  $x \in \Sigma_{\mathcal{I}}$ . Furthermore, in each of the first two expressions, the infimum over all  $n \geq 1$  may be replaced with a limit as  $n \rightarrow \infty$  without affecting the validity of the expression.*

This theorem motivates us to study the following two classes of objects. Firstly, given a compact set  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  indexed in  $\mathcal{I}$ , let us define the set of *maximising measures* of  $\mathbf{A}$  to be the set of measures on  $\Sigma_{\mathcal{I}}$  given by

$$\mathcal{M}_{\max}(\mathbf{A}) := \left\{ \mu \in \mathcal{M}_{\sigma} : \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) = \log \varrho(\mathbf{A}) \right\}.$$

Secondly, we shall say that  $x \in \Sigma_{\mathcal{I}}$  is *strongly extremal* if there exists  $\varepsilon > 0$  such that  $\|\mathcal{L}_{\mathbf{A}}(x, n)\| \geq \varepsilon \varrho(\mathbf{A})^n$  for every  $n \geq 1$ , and  $x$  is *weakly extremal* if  $\lim_{n \rightarrow \infty} \|\mathcal{L}_{\mathbf{A}}(x, n)\|^{1/n} = \varrho(\mathbf{A})$ . These definitions were previously given in a more specialised context in [25]; related definitions have been applied independently by other authors [33, 57], and some comparison with these definitions is undertaken in §6. Since every norm on  $\mathbf{M}_d(\mathbb{C})$  is equivalent to  $\|\cdot\|$ , the sets of strongly and weakly extremal elements of  $\Sigma_{\mathcal{I}}$  are unaffected if a different norm is used in the definition.

The elementary properties of  $\mathcal{M}_{\max}(\mathbf{A})$  are given by the following:

**Proposition 2.2.** *Let  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  be a compact set indexed in  $\mathcal{I}$ , and suppose that  $\varrho(\mathbf{A}) > 0$ . Then  $\mathcal{M}_{\max}(\mathbf{A})$  is compact, convex and nonempty, and its extreme points are precisely its ergodic elements. If  $\mu \in \mathcal{M}_{\sigma}$  and  $\|\cdot\|$  is any norm on  $\mathbf{M}_d(\mathbb{C})$ , then  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  if and only if  $\lim_{n \rightarrow \infty} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) =$*

$\log \varrho(\mathbf{A})$ . If in addition  $\|\cdot\|$  is submultiplicative, then  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  if and only if  $\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) = \log \varrho(\mathbf{A})$ .

*A priori*, if  $\mu$  is an ergodic maximising measure for  $\mathbf{A}$ , then by the subadditive ergodic theorem it follows that almost every point in the support of  $\mu$  is weakly extremal. By Theorem 2.1, we also know that strongly extremal orbits for  $\mathbf{A}$  exist. Beyond this, it is not automatically obvious that the maximising measures and extremal orbits associated to a set  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  should enjoy any significant structural relationships. Our main theorem shows that in fact both the maximising measures and the extremal orbits of a set  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  are strongly characterised by a certain compact invariant set  $Z_{\mathbf{A}} \subseteq \Sigma_{\mathcal{I}}$ , which we call the *Mather set* of  $\mathbf{A}$ . This result echoes some fundamental results in ergodic optimisation and Lagrangian dynamics, as described below. This theorem is the following:

**Theorem 2.3.** *Let  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  be a compact set with positive joint spectral radius, and let  $\mathcal{I}$  be an index set for  $\mathbf{A}$ . Define the Mather set  $Z_{\mathbf{A}} \subseteq \Sigma_{\mathcal{I}}$  associated to  $\mathbf{A}$  by*

$$Z_{\mathbf{A}} := \bigcup_{\mu \in \mathcal{M}_{\max}(\mathbf{A})} \text{supp } \mu.$$

*Then  $Z_{\mathbf{A}}$  has the following properties:*

- (1)  *$Z_{\mathbf{A}}$  is equal to the support of some measure  $\hat{\mu} \in \mathcal{M}_{\sigma}$ . In particular,  $Z_{\mathbf{A}}$  is compact and nonempty and satisfies  $\sigma Z_{\mathbf{A}} = Z_{\mathbf{A}}$ .*
- (2) *For every  $\mu \in \mathcal{M}_{\sigma}$  we have  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  if and only if  $\mu(Z_{\mathbf{A}}) = 1$ .*
- (3) *Every  $x \in Z_{\mathbf{A}}$  which is recurrent with respect to  $\sigma$  satisfies the property  $\limsup_{n \rightarrow \infty} \varrho(\mathbf{A})^{-n} \rho(\mathcal{L}_{\mathbf{A}}(x, n)) = 1$ .*
- (4) *Every  $x \in Z_{\mathbf{A}}$  is strongly extremal. If in addition  $\mathbf{A}$  is relatively product bounded, then for every extremal norm  $\|\cdot\|$  for  $\mathbf{A}$  we have  $\|\mathcal{L}_{\mathbf{A}}(x, n)\| = \varrho(\mathbf{A})^n$  for all  $x \in Z_{\mathbf{A}}$  and  $n \geq 1$ .*
- (5) *Let  $d$  be a metric which generates the topology of  $\Sigma_{\mathcal{I}}$ , and define  $\text{dist}(x, Z_{\mathbf{A}}) := \inf\{d(x, y) : y \in Z_{\mathbf{A}}\}$  for every  $x \in \Sigma_{\mathcal{I}}$ . If  $z \in \Sigma_{\mathcal{I}}$  is weakly extremal, then  $\lim_{n \rightarrow \infty} (1/n) \sum_{k=0}^{n-1} \text{dist}(\sigma^k z, Z_{\mathbf{A}}) = 0$ .*

*Remark.* The definition of  $Z_{\mathbf{A}}$  and parts (i) and (ii) of Theorem 2.3 were suggested by well-known results of J. Mather and R. Mañé in Lagrangian dynamics [39, 40]; similar results have been obtained in other dynamical contexts by T. Bousch [11] and by G. Contreras *et al.* [15]. Additionally, for the one-parameter family of pairs of matrices  $\mathbf{A}_{\alpha} = \{A_1^{(\alpha)}, A_2^{(\alpha)}\}$  given by

$$A_1^{(\alpha)} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A_2^{(\alpha)} = \begin{pmatrix} \alpha & 0 \\ \alpha & \alpha \end{pmatrix},$$

where  $\alpha \in [0, 1]$ , it is shown in [25, Theorem 2.3] that there exists for each  $\alpha$  a set  $X_{\tau(\alpha)} \subset \Sigma_2$  which satisfies properties (iii)-(v) above. Theorem 2.3 above can therefore also be seen as a partial generalisation of [25, Theorem 2.3] to arbitrary compact sets of square matrices.

Theorem 2.3 allows us to prove a number of theorems relating to optimal growth of matrix sequences, which are described in detail in subsequent sections of this article. In §5, we prove a theorem which relates two different notions of stability for discrete linear inclusions introduced by L. Gurvits in the influential article [24]. In §6 we prove a proposition which unifies various definitions of the ‘1-ratio’ of a pair of matrices given respectively by V. S. Kozyakin, by T. Bousch and J. Mairesse, and

by K. Hare *et al.* Using this unified definition, we prove a result on the continuity of 1-ratios which generalises results by each of these authors. Connections with the Lagarias-Wang finiteness property are discussed. Finally, in §7 we prove a theorem dealing with the classification of Barabanov norms, and prove uniqueness of these norms for a family of pairs of matrices studied in [7, 25, 57]. We defer detailed descriptions of these results to the relevant sections.

The proofs of Theorem 2.1, Proposition 2.2 and Theorem 2.3 are given in §4, following some preparatory results in §3. The first two results in particular are dependent on general results on subadditive sequences of functions, which are presented in the appendix.

### 3. PRELIMINARIES TO THE PROOF OF THEOREM 2.3

In this section we prove some preliminary results needed for the proof of Theorem 2.3, the purpose of which is to allow us to deal with the case in which  $\mathbf{A}$  is not relatively product bounded. In this and the following section, the symbol  $|L|$  will be used to denote the norm of the matrix  $L$  given by the maximum of the moduli of the entries of  $L$ . This is the only norm which we shall apply to matrices which are strictly rectangular.

We begin with the following key definition. Let  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  be compact, and let  $\mathcal{I}$  be an index set for  $\mathbf{A}$ . If  $\mathbf{A}^{(1)} = \{A_i^{(1)} : i \in \mathcal{I}\}$ ,  $\mathbf{A}^{(2)} = \{A_i^{(2)} : i \in \mathcal{I}\}$  and  $\mathbf{B} = \{B_i : i \in \mathcal{I}\}$  are sets of matrices respectively of dimension  $k \times k$ ,  $(d-k) \times (d-k)$  and  $k \times (d-k)$ , where  $0 < k < d$ , we say that  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, M)$  is an *upper triangularisation* of  $\mathbf{A}$  if the relation

$$M^{-1} A_i M = \begin{pmatrix} A_i^{(1)} & B_i \\ 0 & A_i^{(2)} \end{pmatrix}$$

is satisfied for all  $i \in \mathcal{I}$ . Note that if this is the case then each of the maps  $i \mapsto A_i^{(1)}$ ,  $i \mapsto A^{(2)}$ ,  $i \mapsto B_i$  is continuous. The first of the two main results of this section is the following, which guarantees the existence of upper triangularisations with favourable properties:

**Proposition 3.1.** *Suppose that  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  has nonzero joint spectral radius and is not relatively product bounded. Then there exists an upper triangularisation  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, M)$  of  $\mathbf{A}$  such that  $\varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A}^{(2)}) = \varrho(\mathbf{A})$  and  $\mathbf{A}^{(1)}$  is relatively product bounded.*

In order to prove Proposition 3.1 we require two lemmas. The following lemma is by now well-known to joint spectral radius researchers; it appears to originate in [1]. For a proof we refer the reader to the book of R. Jungers [29].

**Lemma 3.2.** *Suppose that  $\mathbf{A}$  is not relatively product bounded. Then there exists a subspace  $U \subset \mathbb{C}^d$  such that  $0 < \dim U < d$  and  $A_i U \subseteq U$  for all  $A_i \in \mathbf{A}$ .*

The statement of the following lemma was suggested to the author during the reviewing process for the article [44] by an anonymous reviewer. To the best of the author's knowledge it has not previously appeared in print.

**Lemma 3.3.** *Suppose that  $\mathbf{A}$  is not relatively product bounded. Then there exists an upper triangularisation  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, M)$  for  $\mathbf{A}$  such that  $\varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A})$ .*

*Proof.* By Lemma 3.2 there exists a subspace  $U$  of  $\mathbb{C}^d$  such that  $A_i U \subseteq U$  for all  $i \in \mathcal{I}$ , and  $0 < \dim U < d$ . Let  $r$  be the largest possible dimension of such a subspace  $U$ , and fix an invariant subspace  $U$  of that dimension. Let  $v_1, \dots, v_d$  be a basis for  $\mathbb{C}^d$  having the property that  $v_1, \dots, v_r$  is a basis for  $U$ , and let  $M$  be the corresponding change-of-basis matrix. Clearly for each  $i \in \mathcal{I}$  we have

$$M^{-1} A_i M = \begin{pmatrix} A_i^{(1)} & B_i \\ 0 & A_i^{(2)} \end{pmatrix}$$

for some  $r \times r$  matrix  $A_i^{(1)}$ ,  $r \times (d-r)$  matrix  $B_i$ , and  $(d-r) \times (d-r)$  matrix  $A_i^{(2)}$ . Define  $B = \{B_i : i \in \mathcal{I}\}$  and  $\mathbf{A}^{(k)} = \{A_i^{(k)} : i \in \mathcal{I}\}$  for  $k = 1, 2$ , so that  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, B, M)$  is an upper triangularisation of  $\mathbf{A}$ .

We claim that there does not exist a subspace  $V$  of  $\mathbb{C}^{d-r}$  such that  $A_i^{(2)} V \subseteq V$  for all  $i \in \mathcal{I}$ , and  $0 < \dim V < d-r$ . To see this it suffices to note that if such a subspace  $V$  were to exist, then  $U \oplus MV$  would be an invariant subspace for  $\mathbf{A}$  with dimension strictly between  $r$  and  $d$ , which is impossible by the definition of  $r$ . We conclude via Lemma 3.2 that  $\mathbf{A}^{(2)}$  is relatively product bounded.

We may now prove that  $\varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A})$ . For a contradiction let us assume that  $\varrho(\mathbf{A}^{(1)}) < \varrho(\mathbf{A})$ . Since  $\mathbf{A}^{(2)}$  is relatively product bounded with joint spectral radius not greater than  $\varrho(\mathbf{A})$ , and since  $\varrho(\mathbf{A}^{(1)}) < \varrho(\mathbf{A})$ , we may choose constants  $C_1, C_2 > 1$  and  $\theta \in (0, \varrho(\mathbf{A}))$  such that for all  $n \geq 1$  we have  $|A_{i_1}^{(1)} \cdots A_{i_n}^{(1)}| \leq C_1 \theta^n$  and  $|A_{i_1}^{(2)} \cdots A_{i_n}^{(2)}| \leq C_2 \varrho(\mathbf{A})^n$  for every  $(i_1, \dots, i_n) \in \mathcal{I}^n$ . Using the fact that  $B$  is compact, we may choose a constant  $C_3 > 1$  such that if  $L_1 \in \mathbf{M}_r(\mathbb{C})$  and  $L_2 \in \mathbf{M}_{d-r}(\mathbb{C})$  then  $|L_1 B_i L_2| \leq C_3 |L_1| |L_2|$  for all  $i \in \mathcal{I}$ . As such, for every  $n \geq 1$  and  $(i_1, \dots, i_n) \in \mathcal{I}^n$  we have

$$\begin{aligned} |M^{-1} A_{i_1} \cdots A_{i_n} M| &= \left| \begin{pmatrix} A_{i_1}^{(1)} \cdots A_{i_n}^{(1)} & \sum_{k=1}^n A_{i_1}^{(1)} \cdots A_{i_{k-1}}^{(1)} B_{i_k} A_{i_{k+1}}^{(2)} \cdots A_{i_n}^{(2)} \\ 0 & A_{i_1}^{(2)} \cdots A_{i_n}^{(2)} \end{pmatrix} \right| \\ &\leq \max \left\{ |A_{i_1}^{(1)} \cdots A_{i_n}^{(1)}|, |A_{i_1}^{(2)} \cdots A_{i_n}^{(2)}|, \sum_{k=1}^n |A_{i_1}^{(1)} \cdots A_{i_{k-1}}^{(1)} B_{i_k} A_{i_{k+1}}^{(2)} \cdots A_{i_n}^{(2)}| \right\}. \end{aligned}$$

Since

$$\begin{aligned} \sum_{k=1}^n |A_{i_1}^{(1)} \cdots A_{i_{k-1}}^{(1)} B_{i_k} A_{i_{k+1}}^{(2)} \cdots A_{i_n}^{(2)}| &\leq \sum_{k=1}^n C_3 |A_{i_1}^{(1)} \cdots A_{i_{k-1}}^{(1)}| \cdot |A_{i_{k+1}}^{(2)} \cdots A_{i_n}^{(2)}| \\ &\leq \sum_{k=1}^n C_1 C_2 C_3 \theta^{k-1} \varrho(\mathbf{A})^{n-k} \\ &< C_1 C_2 C_3 \varrho(\mathbf{A})^{n-1} \sum_{k=0}^{\infty} \left( \frac{\theta}{\varrho(\mathbf{A})} \right)^k \\ &= \varrho(\mathbf{A})^n \cdot \frac{C_1 C_2 C_3}{\varrho(\mathbf{A}) - \theta} \end{aligned}$$

and

$$\max \left\{ |A_{i_1}^{(1)} \cdots A_{i_n}^{(1)}|, |A_{i_1}^{(2)} \cdots A_{i_n}^{(2)}| \right\} \leq \varrho(\mathbf{A})^n \max\{C_1, C_2\},$$

we conclude that

$$\max \left\{ \varrho(\mathbf{A})^{-n} |M^{-1} A_{i_1} \cdots A_{i_n} M| : (i_1, \dots, i_n) \in \mathcal{I}^n \right\} \leq \max \left\{ \frac{C_1 C_2 C_3}{\varrho(\mathbf{A}) - \theta}, C_1, C_2 \right\}$$

for every  $n \geq 1$ . The invertibility of  $M$  implies that the functional  $A \mapsto |M^{-1}AM|$  is a norm on  $\mathbf{M}_d(\mathbb{C})$ , and hence we have shown that  $\mathbf{A}$  is relatively product bounded. This contradicts the hypotheses of the lemma, and we conclude that necessarily  $\varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A})$ . The proof is complete.  $\square$

of *Proposition 3.1*. Let  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, M)$  be an upper triangularisation of  $\mathbf{A}$  such that  $\varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A})$ , which exists by Lemma 3.3, and suppose that the matrices comprising  $\mathbf{A}^{(1)}$  have the smallest possible dimension  $r$  for which this relationship can hold. It follows in particular that there cannot exist an upper triangularisation  $(\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{D}, \tilde{M})$  of  $\mathbf{A}^{(1)}$  such that  $\varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{C}^{(1)})$ , since this would lead to a new upper triangularisation of  $\mathbf{A}$  in which the upper-left matrices have smaller than the minimum possible dimension. By Lemma 3.3 it follows that  $\mathbf{A}^{(1)}$  must be relatively product bounded.

Now let us suppose for a contradiction that  $\varrho(\mathbf{A}^{(2)}) < \varrho(\mathbf{A})$ . We make estimates similar to those in the proof of Lemma 3.3. Choose a constant  $C_3 > 1$  such that if  $L_1 \in \mathbf{M}_r(\mathbb{C})$ ,  $L_2 \in \mathbf{M}_{d-r}(\mathbb{C})$  and  $i \in \mathcal{I}$  then  $|L_1 B_i L_2| \leq C_3 |L_1| |L_2|$ , and choose constants  $C_1, C_2 > 1$  and  $\theta < \varrho(\mathbf{A})$  such that for each  $n \geq 1$  and  $(i_1, \dots, i_n) \in \mathcal{I}^n$  we have  $|A_{i_1}^{(1)} \cdots A_{i_n}^{(1)}| \leq C_1 \varrho(\mathbf{A})^n$  and  $|A_{i_1}^{(2)} \cdots A_{i_n}^{(2)}| \leq C_2 \theta^n$ . Given any  $n \geq 1$  and  $(i_1, \dots, i_n) \in \mathcal{I}^n$ , the quantity

$$\max \left\{ \left| A_{i_1}^{(1)} \cdots A_{i_n}^{(1)} \right|, \left| A_{i_1}^{(2)} \cdots A_{i_n}^{(2)} \right|, \sum_{k=1}^n \left| A_{i_1}^{(1)} \cdots A_{i_{k-1}}^{(1)} B_{i_k} A_{i_{k+1}}^{(2)} \cdots A_{i_n}^{(2)} \right| \right\}$$

is as before an upper bound for  $|M^{-1} A_{i_1} \cdots A_{i_n} M|$ , and in a similar manner to the previous proof we obtain the estimates

$$\sum_{k=1}^n \left| A_{i_1}^{(1)} \cdots A_{i_{k-1}}^{(1)} B_{i_k} A_{i_{k+1}}^{(2)} \cdots A_{i_n}^{(2)} \right| \leq \sum_{k=1}^n C_1 C_2 C_3 \varrho(\mathbf{A})^{k-1} \theta^{n-k} < \varrho(\mathbf{A})^n \frac{C_1 C_2 C_3}{\varrho(\mathbf{A}) - \theta}$$

and

$$\max \left\{ \left| A_{i_1}^{(1)} \cdots A_{i_n}^{(1)} \right|, \left| A_{i_1}^{(2)} \cdots A_{i_n}^{(2)} \right| \right\} \leq \varrho(\mathbf{A})^n \max\{C_1, C_2\}.$$

It follows as before that the quantity

$$\max \left\{ \varrho(\mathbf{A})^{-n} \|A_{i_1} \cdots A_{i_n}\| : (i_1, \dots, i_n) \in \mathcal{I}^n \right\}$$

is bounded independently of  $n$ , contradicting the hypothesis that  $\mathbf{A}$  is not relatively product bounded. We conclude that necessarily  $\varrho(\mathbf{A}^{(2)}) = \varrho(\mathbf{A})$  as claimed. This completes the proof of the proposition.  $\square$

The second major result of this section is the following, which allows us to apply the properties of the upper triangularisations given by Proposition 3.1 to the comparison of growth rates along certain sequences.

**Proposition 3.4.** *Let  $\mathbf{A}$  be a compact set of matrices indexed in  $\mathcal{I}$ , and suppose that  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, M)$  is an upper triangularisation of  $\mathbf{A}$ . Then for every  $\mu \in \mathcal{E}_\sigma$ ,*

$$\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_\mathbf{A}(x, n)\| d\mu(x) = \max_{j=1,2} \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}^{(j)}}(x, n)\| d\mu(x).$$

In order to prove the proposition we require the following abstract lemma:

**Lemma 3.5.** *Let  $(X, \mathcal{F}, \nu)$  be a probability space and let  $T: X \rightarrow X$  be an ergodic measure-preserving transformation. Suppose that  $(f_n^1), (f_n^2)$  are sequences of measurable functions from  $X$  to  $\mathbb{R} \cup \{-\infty\}$  such that  $f_{n+m}^j \leq f_n^j \circ T^m + f_m^j$   $\nu$ -a.e. for all  $n, m \geq 1$ , for each  $j \in \{1, 2\}$ , and such that each  $f_n^j$  is bounded above by an integrable function. Define  $\lambda_j := \inf_{n \geq 1} \frac{1}{n} \int f_n^j d\nu$  and  $f_0^j \equiv 0$  for each  $j$ . Then for  $\nu$ -a.e.  $x \in X$ ,*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \max_{1 \leq k \leq n} (f_{n-k}^1(T^k x) + f_{k-1}^2(x)) \leq \max\{\lambda_1, \lambda_2\}.$$

*Proof.* By passing to the natural extension of  $(X, T)$  if necessary, we may without loss of generality assume that  $T$  is invertible. We will show that for each integer  $\ell \geq 1$ ,

$$(4) \quad \nu \left( \left\{ x \in X : \liminf_{n \rightarrow \infty} \frac{1}{n} \max_{1 \leq k \leq n} (f_{n-k}^1(T^k x) + f_{k-1}^2(x)) \leq \max\{\lambda_1, \lambda_2\} + \frac{1}{\ell} \right\} \right) = 1,$$

which suffices to prove the lemma.

Applying Theorem A.2 to the sequence  $(f_n^2)$  and the transformation  $T$  we find that  $n^{-1} f_n^2 \rightarrow \lambda_2$   $\nu$ -almost everywhere. On the other hand, our hypotheses imply that  $f_{n+m}^1 \circ T^{-n-m} \leq f_n^1 \circ T^{-n-m} + f_m^1 \circ T^{-m}$  almost everywhere for each  $n, m \geq 1$ , and hence we may apply Theorem A.2 to the sequence  $(f_n^1 \circ T^{-n})$  and the transformation  $T^{-1}$  to deduce that  $f_n^1 \circ T^{-n} \rightarrow \lambda_1$   $\nu$ -a.e. Fix  $\ell \geq 1$ , and for each  $x \in X$  define  $C_\ell^1(x) := \sup\{f_n^1(T^{-n}x) - n(\lambda_1 + \ell^{-1}) : n \geq 0\}$  and  $C_\ell^2(x) := \sup\{f_n^2(x) - n(\lambda_2 + \ell^{-1}) : n \geq 0\}$ . Note that each  $C_\ell^j(x)$  is finite for  $\nu$ -a.e.  $x \in X$ . For each  $x \in X$  and  $n \geq 1$  we then have

$$(5) \quad \begin{aligned} & \max_{1 \leq k \leq n} (f_{n-k}^1(T^k x) + f_{k-1}^2(x)) \\ & \leq \max_{1 \leq k \leq n} (C_\ell^1(T^n x) + (n-k)(\lambda_2 + \ell^{-1}) + C_\ell^2(x) + (k-1)(\lambda_1 + \ell^{-1})) \\ & = C_\ell^1(T^n x) + C_\ell^2(x) + (n-1) \max\{\lambda_1, \lambda_2\} + (n-1)\ell^{-1}. \end{aligned}$$

Now choose some  $K \in \mathbb{R}$  such that  $\nu(\{x \in X : C_\ell^1(x) \leq K\}) > 0$ . By ergodicity we have  $C_\ell^1(T^n x) \leq K$  infinitely often for  $\nu$ -a.e.  $x$ , and it follows that  $\liminf_{n \rightarrow \infty} n^{-1} C_\ell^1(T^n x) \leq 0$   $\nu$ -a.e. Combining this fact with the inequality (5) yields (4), and since  $\ell \geq 1$  is arbitrary the lemma follows.  $\square$

of Proposition 3.4. Let us define

$$\lambda := \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_A(x, n)\| d\mu(x)$$

and

$$\lambda_j := \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{A^{(j)}}(x, n)\| d\mu(x)$$

for  $j = 1, 2$ . From the definition of upper triangularisation, we have for each  $n \geq 1$  and all  $x \in \Sigma_{\mathcal{I}}$ ,

$$(6) \quad |M^{-1} \mathcal{L}_A(x, n) M| = \left| \begin{pmatrix} \mathcal{L}_{A^{(1)}}(x, n) & \sum_{k=1}^n \mathcal{L}_{A^{(1)}}(\sigma^k x, n-k) B_{x_k} \mathcal{L}_{A^{(2)}}(x, k-1) \\ 0 & \mathcal{L}_{A^{(2)}}(x, n) \end{pmatrix} \right|,$$

so that in particular

$$(7) \quad \frac{1}{n} \log |M^{-1} \mathcal{L}_A(x, n) M| \geq \max_{j \in \{1, 2\}} \frac{1}{n} \log |\mathcal{L}_{A^{(j)}}(x, n)|.$$

Since  $M$  is invertible, the functional  $A \mapsto |M^{-1}AM|$  defines a norm on  $\mathbf{M}_d(\mathbb{C})$ , and so we may deduce from (7) the inequality

$$\begin{aligned} \lambda &= \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_A(x, n)\| d\mu(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \int \log \|\mathcal{L}_A(x, n)\| d\mu(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \int \log |M^{-1} \mathcal{L}_A(x, n) M| d\mu(x) \\ &\geq \liminf_{n \rightarrow \infty} \max_{j \in \{1, 2\}} \frac{1}{n} \int \log |\mathcal{L}_{A^{(j)}}(x, n)| d\mu(x) \\ &\geq \max_{j \in \{1, 2\}} \liminf_{n \rightarrow \infty} \frac{1}{n} \int \log \|\mathcal{L}_{A^{(j)}}(x, n)\| d\mu(x) \\ &= \max_{j \in \{1, 2\}} \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{A^{(j)}}(x, n)\| d\mu(x) \\ &= \max\{\lambda_1, \lambda_2\}, \end{aligned}$$

where we have used Lemma A.1 to identify the infima with the corresponding limits. Let us now prove the reverse inequality  $\lambda \leq \max\{\lambda_1, \lambda_2\}$ . As a consequence of (6), the quantity

$$\max \left\{ |\mathcal{L}_{A^{(1)}}(x, n)|, |\mathcal{L}_{A^{(2)}}(x, n)|, \sum_{k=1}^n |\mathcal{L}_{A^{(1)}}(\sigma^k x, n-k) B_{x_k} \mathcal{L}_{A^{(2)}}(x, k-1)| \right\}$$

is an upper bound for  $|M^{-1} \mathcal{L}_A(x, n) M|$  for every  $x$  and  $n$ . Choose  $C > 1$  such that  $|L_1 B_i L_2| \leq C \|L_1\| \|L_2\|$  for all  $i \in \mathcal{I}$ ,  $L_1 \in \mathbf{M}_r(\mathbb{C})$  and  $L_2 \in \mathbf{M}_{d-r}(\mathbb{C})$ . Applying Lemma 3.5 with  $f_n^j(x) := \log \|\mathcal{L}_{A^{(j)}}(x, n)\|$  we obtain for  $\mu$ -a.e.  $x \in \Sigma_{\mathcal{I}}$ ,

$$\begin{aligned} &\liminf_{n \rightarrow \infty} \frac{1}{n} \log \left( \sum_{k=1}^n |\mathcal{L}_{A^{(1)}}(\sigma^k x, n-k) B_{x_k} \mathcal{L}_{A^{(2)}}(x, k-1)| \right) \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \left( C n \max_{1 \leq k \leq n} \|\mathcal{L}_{A^{(1)}}(\sigma^k x, n-k)\| \cdot \|\mathcal{L}_{A^{(2)}}(x, k-1)\| \right) \leq \max\{\lambda_1, \lambda_2\}. \end{aligned}$$

Applying Theorem A.2 to each sequence  $(f_n^j)$  and making use of the equivalence of norms on finite-dimensional spaces we obtain for  $\mu$ -a.e.  $x$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \max_{j \in \{1, 2\}} \log |\mathcal{L}_{A^{(j)}}(x, n)| = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{j \in \{1, 2\}} \log \|\mathcal{L}_{A^{(j)}}(x, n)\| = \max\{\lambda_1, \lambda_2\}.$$

Combining these estimates, it follows that for  $\mu$ -a.e.  $x \in \Sigma_{\mathcal{I}}$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_A(x, n)\| = \liminf_{n \rightarrow \infty} \frac{1}{n} \log |M^{-1} \mathcal{L}_A(x, n) M| \leq \max\{\lambda_1, \lambda_2\}.$$

Applying Theorem A.2 once more with  $f_n(x) := \log \|\mathcal{L}_A(x, n)\|$  we conclude that for  $\mu$ -a.e.  $x$ ,

$$\lambda = \liminf_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_A(x, n)\| \leq \max\{\lambda_1, \lambda_2\},$$

and therefore  $\lambda = \max\{\lambda_1, \lambda_2\}$  as required. The proof is complete.  $\square$

*Remark.* In general the ergodicity hypothesis in Proposition 3.4 may not be removed. We note the following example. Let  $\mathcal{I} = \{1, 2\}$  and define

$$A_1 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

and  $\mathbf{A} := \{A_1, A_2\}$ . If we define sets  $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}$  and  $\mathbf{B}$  of  $1 \times 1$  matrices indexed over  $\mathcal{I}$  by  $A_1^{(1)} = A_2^{(2)} = 3$ ,  $A_2^{(1)} = A_1^{(2)} = 1$  and  $B_1 = B_2 = 0$ , then  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, I)$  defines an upper triangularisation of  $\mathbf{A}$ . Let  $\delta_1$  be the Dirac measure on  $\Sigma_2$  concentrated at the constant sequence  $y = (y_i) \in \Sigma_2$  whose entries are all equal to 1, let  $\delta_2$  be the Dirac measure concentrated on the constant sequence  $z = (z_i) \in \Sigma_2$  whose entries are all equal to 2, and define  $\mu := \frac{1}{2}(\delta_1 + \delta_2)$ . An elementary calculation shows that

$$\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) = \log 3,$$

$$\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}^{(1)}}(x, n)\| d\mu(x) = \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}^{(2)}}(x, n)\| d\mu(x) = \frac{\log 3}{2}$$

so that the conclusion of Proposition 3.4 is not valid in this case.

#### 4. PROOFS OF MAIN THEOREMS

We begin with the proofs of Theorem 2.1 and Proposition 2.2, which are direct consequences of general results on subadditive sequences of functions given in the appendix.

*of Theorem 2.1.* Since  $\|\cdot\|$  and  $\|\cdot\|$  are norms on the finite-dimensional space  $\mathbf{M}_d(\mathbb{C})$  they are equivalent, and so there exists a constant  $C > 1$  such that  $C^{-1}\|B\| \leq \|\cdot\| \leq C\|B\|$  for every  $B \in \mathbf{M}_d(\mathbb{C})$ . By (2) it follows that

$$\log \varrho(\mathbf{A}) = \lim_{n \rightarrow \infty} \sup \left\{ \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| : x \in \Sigma_{\mathcal{I}} \right\} = \lim_{n \rightarrow \infty} \sup \left\{ \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| : x \in \Sigma_{\mathcal{I}} \right\}.$$

Define a sequence of functions  $f_n: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R} \cup \{-\infty\}$  by  $f_n(x) := \log \|\mathcal{L}_{\mathbf{A}}(x, n)\|$  for each  $n \geq 1$  and  $x \in \Sigma_{\mathcal{I}}$ . Applying Theorem A.3 completes the proof of Theorem 2.1.  $\square$

*of Proposition 2.2.* Define a sequence of functions  $f_n: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R} \cup \{-\infty\}$  by  $f_n(x) := \log \|\mathcal{L}_{\mathbf{A}}(x, n)\|$  for each  $n \geq 1$  and  $x \in \Sigma_{\mathcal{I}}$ . By definition  $\mathcal{M}_{\max}(\mathbf{A})$  is precisely the set of all measures  $\mu \in \mathcal{M}_{\sigma}$  such that  $\inf_{n \geq 1} \frac{1}{n} \int f_n d\mu = \log \varrho(\mathbf{A})$ . By Proposition A.5 this set is nonempty, compact and convex, and its extreme points are precisely its ergodic elements. If  $\|\cdot\|$  is any norm on  $\mathbf{M}_d(\mathbb{C})$ , choose a constant  $C > 1$  such that  $C^{-1}\|B\| \leq \|\cdot\| \leq C\|B\|$  for every  $B \in \mathbf{M}_d(\mathbb{C})$ , and define  $g_n(x) := \log \|\mathcal{L}_{\mathbf{A}}(x, n)\|$  for each  $n \geq 1$  and  $x \in \Sigma_{\mathcal{I}}$ . We have  $|f_n(x) - g_n(x)| \leq \log C$  for all  $x$  and  $n$ . If  $\mu \in \mathcal{M}_{\sigma}$  then the sequence  $(a_n)$  given by  $a_n := \int f_n d\mu$  is subadditive, and so using Lemma A.1 we obtain

$$\inf_{n \geq 1} \frac{1}{n} \int f_n d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \int f_n d\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \int g_n d\mu.$$

In particular  $\mu \in \mathcal{M}_{\sigma}$  if and only if  $\lim_{n \rightarrow \infty} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| d\mu(x) = \log \varrho(\mathbf{A})$ . If additionally  $\|\cdot\|$  is submultiplicative then the sequence  $(\int g_n d\mu)$  is subadditive, so by Lemma A.1 we have  $\lim_{n \rightarrow \infty} \frac{1}{n} \int g_n d\mu = \inf_{n \geq 1} \frac{1}{n} \int g_n d\mu$  in this case. The result follows.  $\square$

We now give the much longer proof of Theorem 2.3, starting with the following important special case.

**Lemma 4.1.** *Let  $\mathbf{A}$  be relatively product bounded. Then parts (i)-(iv) of the conclusions of Theorem 2.3 hold for  $\mathbf{A}$ .*

*Proof.* Let  $\mathcal{I}$  be an index set for  $\mathbf{A}$ , and let us denote by  $Z$  the set  $Z_{\mathbf{A}} \subseteq \Sigma_{\mathcal{I}}$  defined in the statement of Theorem 2.3. We begin by proving (i), for which we recycle a simple argument from [42, Proposition 1].

By Proposition 2.2,  $\mathcal{M}_{\max}(\mathbf{A})$  is a compact subset of a metrisable topological space, and therefore it is separable. Let  $(\mu_n)_{n=1}^{\infty}$  be a sequence of measures which is dense in  $\mathcal{M}_{\max}(\mathbf{A})$ , and define  $\hat{\mu} := \sum_{n=1}^{\infty} 2^{-n} \mu_n$ . Since  $\mathcal{M}_{\max}(\mathbf{A})$  is compact and convex we have  $\hat{\mu} \in \mathcal{M}_{\max}(\mathbf{A})$ , and therefore  $\text{supp } \hat{\mu} \subseteq Z$ . Now, if  $U \subset \Sigma_{\mathcal{I}}$  is open with  $\nu(U) > 0$  for some  $\nu \in \mathcal{M}_{\max}(\mathbf{A})$ , then the set  $\mathcal{U} := \{\mu \in \mathcal{M}_{\max}(\mathbf{A}) : \mu(U) > 0\}$  is a nonempty open subset of  $\mathcal{M}_{\max}(\mathbf{A})$ . Consequently there exists  $k \geq 1$  such that  $\mu_k \in \mathcal{U}$ , and therefore  $\hat{\mu}(U) \geq 2^{-k} \mu_k(U) > 0$ . We deduce that the open set  $\Sigma_{\mathcal{I}} \setminus \text{supp } \hat{\mu}$  cannot have positive measure with respect to any  $\nu \in \mathcal{M}_{\max}(\mathbf{A})$ , and it follows that  $Z \subseteq \text{supp } \hat{\mu}$ . We conclude that  $Z = \text{supp } \hat{\mu}$  and part (i) of Theorem 2.3 holds for  $\mathbf{A}$ .

We next prove both (ii) and (iv). Since  $\mathbf{A}$  is assumed to be relatively product bounded, it has at least one extremal norm. Let  $\|\cdot\|$  be any extremal norm for  $\mathbf{A}$ , and for each  $n \geq 1$  and  $x \in \Sigma_{\mathcal{I}}$  define  $f_n(x) := \log \|\mathcal{L}_{\mathbf{A}}(x, n)\|$ . Clearly  $\sup_{x \in \Sigma_{\mathcal{I}}} f_n(x) = n \log \varrho(\mathbf{A})$  for every  $n \geq 1$ . It follows by Lemma A.9 that the set  $Y_{\|\cdot\|} := \{x \in \Sigma_{\mathcal{I}} : \|\mathcal{L}_{\mathbf{A}}(x, n)\| = \varrho(\mathbf{A})^n \forall n \geq 1\}$  is compact and nonempty, satisfies  $\sigma Y_{\|\cdot\|} \subseteq Y_{\|\cdot\|}$ , and has the property that for each measure  $\mu \in \mathcal{M}_{\sigma}$  we have  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  if and only if  $\mu(Y_{\|\cdot\|}) = 1$ . Since by part (i)  $Z$  is equal to the support of  $\hat{\mu} \in \mathcal{M}_{\max}(\mathbf{A})$ , we in particular have  $Z \subseteq Y_{\|\cdot\|}$ . It follows that if  $\mu \in \mathcal{M}_{\sigma}$  and  $\mu(Z) = 1$ , then  $\mu(Y_{\|\cdot\|}) = \mu(Z) = 1$  and therefore  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ , which establishes (ii). Since  $Z \subseteq Y_{\|\cdot\|}$ , and  $\|\cdot\|$  is an arbitrary extremal norm, we immediately deduce (iv).

It remains to prove (iii), for which we modify an argument from [44]. Let  $\|\cdot\|$  be an extremal norm for  $\mathbf{A}$ , and fix any recurrent  $x \in Z$ . Define

$$\mathcal{S}_{\mathbf{A}}(x) := \bigcap_{m=1}^{\infty} \left( \overline{\bigcup_{n=m}^{\infty} \left\{ \varrho(\mathbf{A})^{-n} \mathcal{L}_{\mathbf{A}}(x, n) : d(\sigma^n x, x) < \frac{1}{m} \right\}} \right).$$

That is,  $\mathcal{S}_{\mathbf{A}}(x)$  is the set of all  $B \in \mathbf{M}_d(\mathbb{C})$  which are equal to the limit as  $j \rightarrow \infty$  of  $\varrho(\mathbf{A})^{-n_j} \mathcal{L}_{\mathbf{A}}(x, n_j)$  along some strictly increasing sequence  $(n_j)_{j=1}^{\infty}$  having the property that  $\lim_{j \rightarrow \infty} \sigma^{n_j} x = x$ . From part (iv) of the Theorem it is clear that  $\|B\| = 1$  for every  $B \in \mathcal{S}_{\mathbf{A}}(x)$ , and since  $\mathcal{S}_{\mathbf{A}}(x)$  is closed it is compact. Since  $x$  is recurrent,  $\mathcal{S}_{\mathbf{A}}(x)$  is nonempty as a consequence of the compactness of the unit sphere of  $\mathbf{M}_d(\mathbb{C})$  with respect to  $\|\cdot\|$ . A simple calculation as in [44] shows that  $\mathcal{S}_{\mathbf{A}}(x)$  is a semigroup with respect to composition of its elements, and since  $\mathcal{S}_{\mathbf{A}}(x)$  is compact it follows that it contains an idempotent element  $P$  (see e.g. [27]). The idempotent matrix  $P$  is a projection with nonzero norm and therefore satisfies  $\rho(P) = 1$ . It follows from the definition of  $\mathcal{S}_{\mathbf{A}}(x)$  that there exists a strictly increasing sequence of integers  $(n_j)_{j=1}^{\infty}$  such that  $\varrho(\mathbf{A})^{-n_j} \|\mathcal{L}_{\mathbf{A}}(x, n_j)\| \rightarrow P$ , and since the spectral radius functional  $\rho: \mathbf{M}_d(\mathbb{C}) \rightarrow \mathbb{R}$  is continuous we have  $\lim_{j \rightarrow \infty} \varrho(\mathbf{A})^{-n_j} \rho(\mathcal{L}_{\mathbf{A}}(x, n_j)) = 1$ . Since clearly  $\rho(\mathcal{L}_{\mathbf{A}}(x, n)) \leq \|\mathcal{L}_{\mathbf{A}}(x, n)\| = \varrho(\mathbf{A})^n$  for all  $n \geq 1$  this proves (iii).  $\square$

We may now prove the Theorem in the general case. We shall prove parts (i) to (iv) of Theorem 2.3 by induction on the dimension  $d$ . If  $d = 1$  then it is trivially

true that  $\mathbf{A}$  is relatively product bounded, so the Theorem holds in the case  $d = 1$  by Lemma 4.1.

Let us now assume that conclusions (i) to (iv) of Theorem 2.3 have been established for all compact sets  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  when  $d$  lies in the range  $1 \leq d \leq D$ , for some natural number  $D$ . We shall show that these conclusions also hold for all compact sets  $\mathbf{A} \subset \mathbf{M}_{D+1}(\mathbb{C})$ . Let us therefore consider a fixed compact set  $\mathbf{A} \subset \mathbf{M}_{D+1}(\mathbb{C})$  indexed over  $\mathcal{I}$ . If  $\mathbf{A}$  is relatively product bounded then the result holds by Lemma 4.1, so for the remainder of the proof the induction step we shall assume that  $\mathbf{A}$  is not relatively product bounded.

By Proposition 3.1 there exists an upper triangularisation  $(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{B}, M)$  of  $\mathbf{A}$  such that  $\varrho(\mathbf{A}) = \varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A}^{(2)})$ . Let us write  $Z_i := Z_{\mathbf{A}^{(i)}} \subseteq \Sigma_{\mathcal{I}}$  for  $i = 1, 2$  and  $Z := Z_{\mathbf{A}} \subseteq \Sigma_{\mathcal{I}}$ , and let  $d_1, d_2$  be respectively the dimensions of the matrices comprising  $\mathbf{A}^{(1)}$  and  $\mathbf{A}^{(2)}$ . Since each  $d_i$  is less than or equal to  $D$ , it follows from the induction hypothesis that each  $\mathbf{A}^{(i)}$  satisfies properties (i) to (iv) of Theorem 2.3.

The core of the proof of the induction step is to show that  $Z = Z_1 \cup Z_2$ . We begin by showing that if  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ , then necessarily  $\mu(Z_1 \cup Z_2) = 1$ . Let  $\mathcal{B}$  be the Borel  $\sigma$ -algebra of the compact metrisable topological space  $\mathcal{M}_{\sigma}$ , and choose any  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ . By Proposition 2.1  $\mathcal{M}_{\max}(\mathbf{A})$  is a compact, convex set whose extremal points are precisely its ergodic elements. By applying a suitable version of Choquet's theorem [47, p.14] it follows that there exists a measure  $\mathbb{P}$  on the measurable space  $(\mathcal{M}_{\max}(\mathbf{A}), \mathcal{B})$  such that  $\int g d\mu = \iint g dm d\mathbb{P}(m)$  for every continuous function  $g: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R}$ , and

$$\mathbb{P} \left( \left\{ \nu \in \mathcal{E}_{\sigma}: \inf_{k \geq 1} \frac{1}{k} \int \log \|\mathcal{L}_{\mathbf{A}}(x, k)\| d\nu(x) = \log \varrho(\mathbf{A}) \right\} \right) = 1.$$

By Proposition 3.4 it follows that for  $\mathbb{P}$ -almost-every  $\nu \in \mathcal{M}_{\sigma}$ ,

$$\log \varrho(\mathbf{A}) = \inf_{k \geq 1} \frac{1}{k} \int \log \|\mathcal{L}_{\mathbf{A}}(x, k)\| d\nu(x) = \max_{i \in \{1, 2\}} \inf_{k \geq 1} \frac{1}{k} \int \log \|\mathcal{L}_{\mathbf{A}^{(i)}}(x, k)\| d\nu(x).$$

In particular, since  $\varrho(\mathbf{A}) = \varrho(\mathbf{A}^{(1)}) = \varrho(\mathbf{A}^{(2)})$  we conclude that  $\mathbb{P}$ -almost-every  $\nu \in \mathcal{M}_{\sigma}$  belongs to  $\mathcal{M}_{\max}(\mathbf{A}^{(1)}) \cup \mathcal{M}_{\max}(\mathbf{A}^{(2)})$ . It follows by the definition of  $Z_1$  and  $Z_2$  that  $\nu(Z_1 \cup Z_2) = 1$  for  $\mathbb{P}$ -a.e.  $\nu \in \mathcal{M}_{\sigma}$ . Since  $Z_1 \cup Z_2$  is closed and  $\Sigma_{\mathcal{I}}$  is metrisable, we may choose a non-negative continuous function  $g: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R}$  which is zero precisely on  $Z_1 \cup Z_2$ . We then have  $\int g d\nu = 0$  for  $\mathbb{P}$ -almost-every  $\nu \in \mathcal{M}_{\sigma}$ , and it follows by integration that  $\int g d\mu = \iint g dm d\mathbb{P}(m) = 0$ . We conclude that  $\mu(Z_1 \cup Z_2) = 1$  as claimed. Since  $\mu$  is an arbitrary element of  $\mathcal{M}_{\max}(\mathbf{A})$ , we furthermore obtain the result that  $Z \subseteq Z_1 \cup Z_2$ .

We now conversely claim that if  $\mu \in \mathcal{M}_{\sigma}$  and  $\mu(Z_1 \cup Z_2) = 1$ , then  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ . Let us fix such a measure  $\mu$ . It is a classical result that the set  $\mathcal{M}_{\sigma}$  is compact and convex, with its set of extremal points being precisely equal to  $\mathcal{E}_{\sigma}$  (see e.g. [59, p.152]). Applying Choquet's theorem we find that there exists a probability measure  $\mathbb{P}$  on the measurable space  $(\mathcal{M}_{\sigma}, \mathcal{B})$  such that  $\int g d\mu = \iint g dm d\mathbb{P}(m)$  for every continuous  $g: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R}$ , and  $\mathbb{P}(\mathcal{E}_{\sigma}) = 1$ . Considering again a non-negative continuous function  $g$  which is zero precisely on  $Z_1 \cup Z_2$ , we note that since  $\int g d\mu = 0$  we must have  $\int g d\nu = 0$  for  $\mathbb{P}$ -a.e.  $\nu \in \mathcal{M}_{\sigma}$ , and therefore  $\nu(Z_1 \cup Z_2) = 1$  for  $\mathbb{P}$ -almost-every  $\nu$ . Since the sets  $Z_1$  and  $Z_2$  are invariant, an ergodic measure which gives full measure to  $Z_1 \cup Z_2$  must also give full measure to at least one of  $Z_1$  and  $Z_2$ . By the induction hypothesis part (ii) of Theorem 2.3 applies to

each  $\mathbf{A}^{(i)}$ , and we deduce that  $\mathbb{P}$ -almost-every measure  $\nu \in \mathcal{M}_\sigma$  belongs to at least one of  $\mathcal{M}_{\max}(\mathbf{A}^{(1)})$  and  $\mathcal{M}_{\max}(\mathbf{A}^{(2)})$ . Applying Proposition 3.4 once more we deduce that  $\mathbb{P}(\mathcal{M}_{\max}(\mathbf{A})) = 1$ . For each fixed pair of integers  $n, k \geq 1$ , the function  $x \mapsto \max\{\log \|\mathcal{L}_\mathbf{A}(x, n)\|, -k\}$  is a continuous function from  $\Sigma_{\mathcal{I}}$  to  $\mathbb{R}$ , and therefore

$$\begin{aligned} \frac{1}{n} \int \max\{\log \|\mathcal{L}_\mathbf{A}(x, n)\|, -k\} d\mu(x) &= \frac{1}{n} \iint \max\{\log \|\mathcal{L}_\mathbf{A}(x, n)\|, -k\} d\nu(x) d\mathbb{P}(\nu) \\ &\geq \frac{1}{n} \iint \log \|\mathcal{L}_\mathbf{A}(x, n)\| d\nu(x) d\mathbb{P}(\nu) \geq \log \varrho(\mathbf{A}). \end{aligned}$$

Letting  $k \rightarrow \infty$  for each fixed  $n$ , applying the monotone convergence theorem and then taking the infimum over  $n$  we obtain  $\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_\mathbf{A}(x, n)\| d\mu(x) \geq \log \varrho(\mathbf{A})$ , and we conclude that  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  as claimed.

We can now show that  $Z = Z_1 \cup Z_2$ . By the induction hypothesis, property (i) of the Theorem holds for  $\mathbf{A}^{(1)}$  and  $\mathbf{A}^{(2)}$ , and hence for  $i = 1, 2$  there exists  $\hat{\mu}_i \in \mathcal{M}_{\max}(\mathbf{A}^{(i)})$  such that  $\text{supp } \hat{\mu}_i = Z_i$ . The  $\sigma$ -invariant measure  $\hat{\mu} := \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)$  has support equal to  $Z_1 \cup Z_2$ , and as previously noted this implies that  $\hat{\mu} \in \mathcal{M}_{\max}(\mathbf{A})$ . We deduce that  $Z_1 \cup Z_2 \subseteq Z$  and therefore  $Z = Z_1 \cup Z_2$  as claimed. Combining this observation with the results already established, it follows immediately that parts (i) and (ii) of Theorem 2.3 hold for  $\mathbf{A}$ .

To complete the proof of the induction step it remains to show that parts (iii) and (iv) of the Theorem hold for  $\mathbf{A}$ . We begin with part (iii). It is a straightforward consequence of the relation

$$M^{-1} \mathcal{L}_\mathbf{A}(x, n) M = \begin{pmatrix} \mathcal{L}_{\mathbf{A}^{(1)}}(x, n) & * \\ 0 & \mathcal{L}_{\mathbf{A}^{(2)}}(x, n) \end{pmatrix}$$

together with Gelfand's formula that  $\rho(\mathcal{L}_\mathbf{A}(x, n)) = \max_{i=1,2} \rho(\mathcal{L}_{\mathbf{A}^{(i)}}(x, n))$  for all  $x \in \Sigma_{\mathcal{I}}$  and  $n \geq 1$ . By the induction hypothesis we know that for each  $i$ , every  $x \in Z_i$  satisfies  $\limsup_{n \rightarrow \infty} \varrho(\mathbf{A}^{(i)})^{-n} \rho(\mathcal{L}_{\mathbf{A}^{(i)}}(x, n)) = 1$ , and since  $Z = Z_1 \cup Z_2$  it follows directly that  $\limsup_{n \rightarrow \infty} \varrho(\mathbf{A})^{-n} \rho(\mathcal{L}_\mathbf{A}(x, n)) = 1$  for every  $x \in Z$ . Similarly, by the induction hypothesis every  $x \in Z_1 \cup Z_2 = Z$  satisfies  $\inf_{n \geq 1} \max_{i=1,2} \varrho(\mathbf{A})^{-n} |\mathcal{L}_{\mathbf{A}^{(i)}}(x, n)| > 0$ , and it follows that every  $x \in Z$  satisfies  $\inf_{n \geq 1} \varrho(\mathbf{A})^{-n} \|\mathcal{L}_\mathbf{A}(x, n)\| > 0$  as required to prove (iv). We have shown that if parts (i)-(iv) of Theorem 2.3 hold for all compact sets  $\mathbf{A}$  of matrices of dimension at most  $D$ , then they also hold for all compact sets of matrices of dimension at most  $D+1$ . It follows by induction that properties (i)-(iv) of Theorem 2.3 hold for every  $\mathbf{A}$  within the scope of the statement of the Theorem.

To complete the proof of the Theorem it remains only to establish (v). Let us fix  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  indexed in  $\mathcal{I}$ , and let  $x \in \Sigma_{\mathcal{I}}$  be weakly extremal. For each  $n \geq 1$  define  $\mu_n := (1/n) \sum_{k=0}^{n-1} \delta_{\sigma^k x} \in \mathcal{M}$ . Applying Lemma A.6 with  $f_n(x) := \|\mathcal{L}_\mathbf{A}(x, n)\|$  we see that every accumulation point of  $(\mu_n)$  lies in  $\mathcal{M}_{\max}(\mathbf{A})$ . Define a continuous function  $g: \Sigma_{\mathcal{I}} \rightarrow \mathbb{R}$  by  $g(x) := \text{dist}(x, Z_\mathbf{A})$ . If  $(\mu_{n_j})$  is a convergent subsequence of  $(\mu_n)$  with limit  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  then clearly  $\int g d\mu = 0$  using the definition of  $Z_\mathbf{A}$ . It follows that  $(1/n_j) \sum_{k=0}^{n_j-1} \text{dist}(\sigma^k x, Z_\mathbf{A}) \rightarrow 0$  along any subsequence  $(n_j)$  such that the sequence of measures  $(\mu_{n_j})$  is convergent in  $\mathcal{M}$ . Since  $\mathcal{M}$  is compact and metrisable, it follows that every subsequence  $(n_j)$  of the natural numbers possesses a finer subsequence with this property, and we conclude that  $(1/n) \sum_{k=0}^{n-1} \text{dist}(\sigma^k x, Z_\mathbf{A}) \rightarrow 0$  as required. The proof of the Theorem is complete.

## 5. APPLICATIONS TO STABILITY THEORY FOR DISCRETE LINEAR INCLUSIONS

Given a finite set  $\mathbf{A} = \{A_1, \dots, A_\ell\} \subset \mathbf{M}_d(\mathbb{C})$ , the discrete linear inclusion associated to  $\mathbf{A}$  is the set of all possible sequences of vectors  $(v_n)_{n \geq 0}$  belonging to  $\mathbb{C}^d$  having the form  $v_n \equiv \mathcal{L}_{\mathbf{A}}(x, n)v_0$  for some  $x \in \Sigma_\ell$ . In the past two decades discrete linear inclusions have attracted substantial interest in the control theory literature [1, 2, 16, 20, 24, 54, 56, 60].

In the influential article [24], L. Gurvits introduced the following notions of stability for discrete linear inclusions. The discrete linear inclusion corresponding to  $\mathbf{A}$  is said to be *absolutely asymptotically stable* if  $v_n \rightarrow 0$  for every trajectory of the inclusion, or equivalently if  $\mathcal{L}_{\mathbf{A}}(x, n) \rightarrow 0$  for every  $x \in \Sigma_\ell$ . We say that the inclusion associated to  $\mathbf{A}$  is *periodically asymptotically stable* if  $\mathcal{L}_{\mathbf{A}}(x, n)v_0 \rightarrow 0$  for all  $v_0 \in \mathbb{C}^d$  whenever  $x \in \Sigma_\ell$  is a periodic sequence, or equivalently if  $\mathcal{L}_{\mathbf{A}}(x, n) \rightarrow 0$  for every periodic  $x \in \Sigma_\ell$ . Finally, the inclusion associated to  $\mathbf{A}$  is *Markov asymptotically stable* if for every Markov measure  $\mu \in \mathcal{M}$  for which all transition probabilities are nonzero, we have  $\mathcal{L}_{\mathbf{A}}(x, n)v_0 \rightarrow 0$   $\mu$ -a.e. for all  $v_0 \in \mathbb{C}^d$ , or equivalently  $\mathcal{L}_{\mathbf{A}}(x, n) \rightarrow 0$   $\mu$ -a.e. In the remainder of this section we shall simply say that  $\mu$  is a *full Markov measure* if it is a Markov measure on  $\Sigma_\ell$  such that all transition probabilities in the associated transition matrix are nonzero. Note that every full Markov measure belongs to  $\mathcal{E}_\sigma$ .

It is not difficult to see that the linear inclusion associated to  $\mathbf{A}$  is absolutely asymptotically stable if and only if  $\varrho(\mathbf{A}) < 1$ , and moreover that absolute asymptotic stability implies both periodic and Markov asymptotic stability. L. Gurvits asked in [24] whether periodic asymptotic stability implies absolute asymptotic stability; the existence of counterexamples to the Lagarias-Wang finiteness conjecture (see section 6 below) answered this question in the negative. On the other hand it is straightforward to exhibit sets of matrices whose associated discrete linear inclusions are Markov asymptotically stable but not absolutely asymptotically stable, for example the set

$$\mathbf{A} := \left\{ \begin{pmatrix} 0 & \lambda \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ \lambda & 0 \end{pmatrix} \right\}$$

has this property for any  $\lambda \geq 1$ . In this section we give a proof of the following result, which was recently announced by X. Dai, Y. Huang and M. Xiao in conference proceedings [17]:

**Theorem 5.1.** *Let  $\mathbf{A} = \{A_1, \dots, A_\ell\} \subset \mathbf{M}_d(\mathbb{C})$ , and suppose that the associated discrete linear inclusion is periodically asymptotically stable. Then it is also Markov asymptotically stable.*

We shall deduce Theorem 5.1 from the following simple lemma:

**Lemma 5.2.** *Let  $\mathbf{A} = \{A_1, \dots, A_\ell\} \subset \mathbf{M}_d(\mathbb{C})$  be a finite set of matrices such that the associated discrete linear inclusion is not Markov asymptotically stable. Then either  $\varrho(\mathbf{A}) > 1$ , or  $\varrho(\mathbf{A}) = 1$  and  $Z_{\mathbf{A}} = \Sigma_\ell$ .*

*Proof.* If  $\mu \in \mathcal{E}_\sigma$  is any full Markov measure, then by the subadditive ergodic theorem (given in the appendix as Theorem A.2) we have for  $\mu$ -almost-every  $x \in \Sigma_\ell$

$$(8) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(x, n)\| = \inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_{\mathbf{A}}(z, n)\| d\mu(z) \leq \log \varrho(\mathbf{A}).$$

If the discrete linear inclusion associated to  $\mathbf{A}$  is not Markov asymptotically stable then there exists a full Markov measure  $\mu \in \mathcal{E}_\sigma$  such that  $\|\mathcal{L}_{\mathbf{A}}(x, n)\|$  does

not converge to zero  $\mu$ -a.e. in the limit as  $n \rightarrow \infty$ . Using (8) we deduce that  $\inf_{n \geq 1} \frac{1}{n} \int \log \|\mathcal{L}_A(x, n)\| d\mu(x) \geq 0$  and therefore  $\varrho(A) \geq 1$ . If  $\varrho(A) > 1$  then there is nothing left to prove, so we shall assume that  $\varrho(A) = 1$ . In this case the measure  $\mu$  belongs to  $\mathcal{M}_{\max}(A)$ , and hence  $\text{supp } \mu$  is contained in  $Z_A$ . Since every full Markov measure has support equal to  $\Sigma_\ell$  we conclude that  $Z_A = \Sigma_\ell$  as claimed.  $\square$

of Theorem 5.1. Suppose that the discrete linear inclusion associated to  $A$  is periodically asymptotically stable. We may deduce from Theorem 2.3(iii) that

$$\varrho(A) = \sup_{n \geq 1} \left\{ \rho(A_{i_1} \cdots A_{i_n})^{\frac{1}{n}} : (i_1, \dots, i_n) \in \{1, \dots, \ell\}^n \right\},$$

an identity originally due to M. A. Berger and Y. Wang [3]. It follows from this identity together with periodic asymptotic stability that  $\varrho(A) \leq 1$ . If  $\varrho(A) < 1$  then the associated discrete linear inclusion is absolutely asymptotically stable and hence is Markov asymptotically stable. If on the other hand  $\varrho(A) = 1$ , then it follows from Theorem 2.3(iv) together with periodic asymptotic stability that the Mather set  $Z_A$  does not contain any periodic orbits. In particular  $Z_A \neq \Sigma_\ell$ , and it follows by Lemma 5.2 that the discrete linear inclusion associated to  $A$  is Markov asymptotically stable. The proof is complete.  $\square$

## 6. APPLICATIONS TO THE STUDY OF THE 1-RATIO OF PAIRS OF MATRICES

In the influential article [37], J. Lagarias and Y. Wang posed the following question: if  $A = \{A_1, \dots, A_\ell\}$  is a finite set of real  $d \times d$  matrices, is it always the case that there exist an integer  $n \geq 1$  and indices  $i_1, \dots, i_n \in \{1, \dots, \ell\}$  such that  $\rho(A_{i_1} \cdots A_{i_n}) = \varrho(A)^n$ ? This question was answered negatively in 2002 by T. Bousch and J. Mairesse [13], but continues to stimulate research [7, 14, 25, 30, 33, 38].

Let us say that  $A$  has the *finiteness property* if Lagarias and Wang's condition holds for  $A$ . Examples of pairs of  $2 \times 2$  matrices which do not satisfy the finiteness property were constructed in [7, 13, 25, 33] by a variety of different methods; the latter three of these four articles possess some important common features which we now describe. In each case one studies a family of pairs of  $2 \times 2$  matrices  $A_\lambda = \{A_1(\lambda), A_2(\lambda)\}$  depending continuously on a parameter  $\lambda$  which belongs to a connected subregion  $\Delta$  of some  $\mathbb{R}^k$ . It is then shown that for each  $\lambda$  in this region, if  $x \in \Sigma_2$  has the property that  $\mathcal{L}_{A_\lambda}(x, n)$  grows rapidly in some suitable sense as  $n \rightarrow \infty$ , then the sequence  $x$  has a well-defined proportion of 1's depending only on  $\lambda$ : that is, in the limit as  $n \rightarrow \infty$  the quantity  $\frac{1}{n} \#\{1 \leq j \leq n : x_j = 1\}$  converges to some value  $\tau(\lambda) \in [0, 1]$ . It is then shown that the 1-ratio function  $\tau: \Delta \rightarrow [0, 1]$  is continuous and nonconstant. Since  $\Delta$  is connected this allows us to deduce that there exists  $\lambda_0 \in \Delta$  such that  $\tau(\lambda_0)$  is irrational. It follows that  $A_{\lambda_0}$  cannot have the finiteness property, since if it were the case that  $\rho(A_{i_m}(\lambda_0) \cdots A_{i_1}(\lambda_0)) = \varrho(A_{\lambda_0})^m$ , then the periodic sequence  $x \in \Sigma_2$  defined by  $x_{km+j} := i_j$  for all  $k \geq 0$  and  $1 \leq j \leq m$  would have the property that  $\mathcal{L}_{A_{\lambda_0}}(x, n)$  grows rapidly, and therefore  $\lim_{n \rightarrow \infty} \frac{1}{n} \#\{1 \leq j \leq n : x_j = 1\} = \tau(\lambda_0) \notin \mathbb{Q}$ ; but by periodicity  $\tau(\lambda_0)$  must be rational with denominator not greater than  $m$ , yielding a contradiction.

We remark that the counterexample to the Lagarias-Wang finiteness conjecture given by Blondel, Theys and Vladimirov [7] is currently the only counterexample in the literature which does not make direct use of the approach given above. However, a detailed investigation by J. Theys [57] shows that a continuous 1-ratio also exists

for the family of pairs of matrices used in [7] to construct a counterexample. This was exploited by Hare, Sidorov, Theys and the present author in [25] to exhibit an *explicit* pair of matrices which fails to have the finiteness property.

A major point of difference between the articles [13, 25, 33] is the precise definition which is used to express the property that  $\mathcal{L}_A(x, n)$  “grows rapidly”. In the examples considered by Bousch and Mairesse, it is shown that for each of the pairs of matrices  $A$  being considered, there is a unique Borel probability measure on  $\Sigma_2$  which is invariant under the shift and which maximises the a.e. pointwise limit of  $\frac{1}{n} \log \|\mathcal{L}_A(x, n)\|$ ; in our formalism this corresponds to showing that  $\mathcal{M}_{\max}(A)$  contains a unique measure. The concept of rapidly growing sequence therefore corresponds to that of a measure-theoretically typical sequence with respect to any maximising measure. In the article [25], a sequence is considered to be “rapidly growing” for the purpose of defining the 1-ratio if it is weakly extremal in the sense defined in §2. Finally, in the article [33], the notion of rapidly growing sequence  $x \in \Sigma_2$  used by Kozyakin is that there exists a Barabanov norm  $\|\cdot\|$  for  $A$  and a vector  $v \in \mathbb{C}^d$  such that  $\varrho(A)^{-n} \|\mathcal{L}_A(x, n)v\| = \|v\| \neq 0$  for every  $n \geq 1$ . One of the objectives of this section is to show that the different definitions of 1-ratio described above yield equivalent results; this is undertaken in Proposition 6.1. We thus give a general and unified definition of the 1-ratio of a finite set of square matrices.

A second point in common between the articles [13, 25, 33] is that the methods used to prove the continuity of the function  $\tau$  do not easily generalise to broader classes of matrices. While the ergodic-theoretic approach of Bousch and Mairesse is relatively general, the context of their theorems imposes strict positivity hypotheses on the matrices comprising the set  $A$ . The arguments in [25] and [33] are much more highly specialised, and seem unlikely to be applicable beyond the case of pairs of real  $2 \times 2$  matrices, with one matrix being upper triangular and the other lower triangular. The second objective of this section is to show in great generality that if a well-defined 1-ratio exists for each member of a family of finite sets of square matrices, then it of necessity varies continuously within that family.

Given integers  $i, \ell$  with  $1 \leq i \leq \ell$ , we let  $[i]$  denote the set of all sequences  $x \in \Sigma_\ell$  whose first entry is equal to  $i$ . For every  $x \in \Sigma_\ell$  and  $n \geq 1$  we thus have  $\sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) = \#\{1 \leq j \leq n: x_j = i\}$ . Note that  $[i]$  is both open and closed as a subset of  $\Sigma_\ell$ . Given a set  $A = \{A_1, \dots, A_\ell\} \subset \mathbf{M}_d(\mathbb{C})$ , let us say that  $x \in \Sigma_\ell$  is *Kozyakin extremal* if there exist a Barabanov norm  $\|\cdot\|$  for  $A$  and a vector  $v \in \mathbb{C}^d$  with  $\|v\| = 1$  such that  $\|\mathcal{L}_A(x, n)v\| = \varrho(A)^n$  for all  $n \geq 1$ . The following proposition shows that the definitions of 1-ratio used by Bousch-Mairesse in [13], by Hare-Morris-Sidorov-Theys in [25] and by Kozyakin in [33] are equivalent:

**Proposition 6.1.** *Let  $A = \{A_1, \dots, A_\ell\} \subset \mathbf{M}_d(\mathbb{C})$ , and let  $\gamma \in [0, 1]$  and  $1 \leq i \leq \ell$ . Then the following properties are equivalent:*

- (1) *For every weakly extremal orbit  $x \in \Sigma_\ell$  we have  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$ .*
- (2) *For every strongly extremal orbit  $x \in \Sigma_\ell$  we have  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$ .*
- (3) *For every ergodic  $\mu \in \mathcal{M}_{\max}(A)$  we have  $\mu([i]) = \gamma$ .*
- (4) *For every  $\mu \in \mathcal{M}_{\max}(A)$  we have  $\mu([i]) = \gamma$ .*

*Suppose additionally that there exists a Barabanov norm for  $A$ . Then each of properties (i)-(iv) is additionally equivalent to the following property: for every Kozyakin extremal  $x \in \Sigma_\ell$ , we have  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$ .*

*Proof.* Since every strongly extremal orbit is weakly extremal it is clear that (i) implies (ii). If  $\mathbf{A}$  satisfies (ii), then by part (iv) of Theorem 2.3 it follows that for every  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$  we have  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) = \gamma$  for  $\mu$ -a.e.  $x$ , and hence by the Birkhoff ergodic theorem we have  $\mu([i]) = \gamma$  for every ergodic  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ , which gives (iii).

If (iii) holds, let us choose any  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ . Using Proposition 2.2 and Choquet's theorem as in previous sections, we may find a measure  $\mathbb{P}$  on the compact convex set  $\mathcal{M}_{\max}(\mathbf{A})$  such that  $\mathbb{P}(\mathcal{M}_{\max}(\mathbf{A}) \cap \mathcal{E}_\sigma) = 1$  and  $\int g d\mu = \int \int g dm d\mathbb{P}(m)$  for all continuous functions  $g: \Sigma_\ell \rightarrow \mathbb{R}$ . The function  $\chi_{[i]}$  is continuous, and since by hypothesis  $\int \chi_{[i]} dm = \gamma$  for  $\mathbb{P}$ -almost-every  $m$  we have  $\mu([i]) = \gamma$  by integration. Since  $\mu$  is arbitrary we conclude that (iv) holds.

Now let us suppose that  $\mathbf{A}$  satisfies (iv). Let  $x \in \Sigma_\ell$  be weakly extremal, and for each  $n \geq 1$  define a Borel probability measure on  $\Sigma_\ell$  by  $\mu_n := \frac{1}{n} \sum_{k=0}^{n-1} \delta_{\sigma^k x}$ , where  $\delta_z$  denotes the Dirac probability measure concentrated at  $z$ . Choose any increasing sequence of integers  $(n_j)$  with the property that  $\frac{1}{n_j} \sum_{k=0}^{n_j-1} \chi_{[i]}(\sigma^k x)$  converges to some limit; we will show that the only possible value of this limit is  $\gamma$ , which implies (i). By replacing  $(n_j)$  with a subsequence if necessary, using the fact that  $\mathcal{M}$  is compact and metrisable, we may assume that  $\mu_{n_j} \rightarrow \mu$  for some measure  $\mu \in \mathcal{M}$ . Since  $x$  is weakly extremal, we may apply Lemma A.6 with  $f_n(x) := \log \|\mathcal{L}_\mathbf{A}(x, n)\|$  to deduce that  $\mu \in \mathcal{M}_\sigma$  and  $\inf_{m \geq 1} \frac{1}{m} \int \log \|\mathcal{L}_\mathbf{A}(x, m)\| d\mu(x) = \log \varrho(\mathbf{A})$  and therefore  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ . Since  $\mu_{n_j} \rightarrow \mu$  and the function  $\chi_{[i]}: \Sigma_\ell \rightarrow \mathbb{R}$  is continuous, it follows that  $\lim_{j \rightarrow \infty} \frac{1}{n_j} \sum_{k=0}^{n_j-1} \chi_{[i]}(\sigma^k x) = \lim_{n \rightarrow \infty} \int \chi_{[i]} d\mu_n = \int \chi_{[i]} d\mu = \mu([i]) = \gamma$ . This completes the proof of the equivalence of statements (i)-(iv).

Finally, let us suppose that a Barabanov norm for  $\mathbf{A}$  exists. We claim that every  $x \in Z_\mathbf{A}$  is Kozyakin extremal. Choose any Barabanov norm  $\|\cdot\|$  for  $\mathbf{A}$ , and let  $x \in Z_\mathbf{A}$ . By part (iv) of Theorem 2.3 we have  $\|\mathcal{L}_\mathbf{A}(x, n)\| = \varrho(\mathbf{A})^n$  for all  $n \geq 1$ . It follows that there exists a subsequence  $(n_j)$  of the natural numbers such that  $\varrho(\mathbf{A})^{-n_j} \mathcal{L}_\mathbf{A}(x, n_j)$  converges as  $j \rightarrow \infty$  to some matrix  $L$  such that  $\|L\| = 1$ . Choose a vector  $v \in \mathbb{C}^d$  such that  $\|v\| = \|Lv\| = 1$ . Fix any  $n \geq 1$  and let  $j$  be sufficiently large that  $n_j > n$ . We have

$$\varrho(\mathbf{A})^{-n_j} \|\mathcal{L}_\mathbf{A}(x, n_j)v\| \leq \varrho(\mathbf{A})^{-n_j} \|\mathcal{L}_\mathbf{A}(\sigma^n x, n_j - n)\| \cdot \|\mathcal{L}_\mathbf{A}(x, n)v\| \leq \varrho(\mathbf{A})^{-n} \|\mathcal{L}_\mathbf{A}(x, n)v\| \leq 1,$$

and since  $\lim_{j \rightarrow \infty} \varrho(\mathbf{A})^{-n_j} \|\mathcal{L}_\mathbf{A}(x, n_j)v\| = \|Lv\| = 1$  we deduce that  $\|\mathcal{L}_\mathbf{A}(x, n)v\| = \varrho(\mathbf{A})^n$ . It follows that  $x$  is Kozyakin extremal as claimed. Now, if (ii) holds, then since every Kozyakin extremal point is strongly extremal, we have  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$  for every Kozyakin extremal  $x \in \Sigma_\ell$ . Conversely, if  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$  for every  $x \in \Sigma_\ell$  which is Kozyakin extremal, then in particular  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$  for every  $x \in Z_\mathbf{A}$ . It follows that  $\frac{1}{n} \sum_{k=0}^{n-1} \chi_{[i]}(\sigma^k x) \rightarrow \gamma$   $\mu$ -a.e. for every ergodic  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ , and by the Birkhoff ergodic theorem this implies that (iii) holds. This completes the proof of the proposition.  $\square$

Given a finite set  $\mathbf{A} = \{A_1, \dots, A_\ell\} \subset \mathbf{M}_d(\mathbb{C})$  and an integer  $1 \leq i \leq \ell$ , we shall say that  $\mathbf{A}$  has a *unique optimal  $i$ -ratio* if there exists a real number  $\gamma \in [0, 1]$  such that one of the conditions in Proposition 6.1 is satisfied. As previously mentioned, some examples of pairs of matrices for which a unique optimal 1-ratio exists are given in [13, 25, 33]. Note that Proposition 6.1 implies that for any  $\ell \geq 2$ , a sufficient condition for  $\mathbf{A} = \{A_1, \dots, A_\ell\}$  to have a unique optimal 1-ratio is that

$\mathcal{M}_{\max}(\mathbf{A})$  contains a unique measure. Conversely, if  $\ell \geq 2$  and  $Z_{\mathbf{A}} = \Sigma_{\ell}$  then it is clear that  $\mathbf{A}$  does not have a unique optimal  $i$ -ratio for any  $i$ ; this situation arises for example if  $\mathbf{A}$  consists entirely of isometries of  $\mathbb{R}^d$ . It is interesting to ask whether a ‘typical’ finite set of  $d \times d$  complex matrices admits a unique optimal 1-ratio. In view of Proposition 6.1, it would be sufficient to show that for a typical finite set of complex  $d \times d$  matrices  $\mathbf{A}$ , the set  $\mathcal{M}_{\max}(\mathbf{A})$  is a singleton set.

The following general result shows that when  $i$ -ratios exist, they are continuous:

**Theorem 6.2.** *Let  $\Delta$  be a metric space, let  $A_1, \dots, A_{\ell}: \Delta \rightarrow \mathbf{M}_d(\mathbb{C})$  be continuous functions, and let  $1 \leq i \leq \ell$ . Suppose that for each  $\lambda \in \Delta$  the set  $\mathbf{A}_{\lambda} := \{A_1(\lambda), \dots, A_{\ell}(\lambda)\}$  has a unique optimal  $i$ -ratio, which we denote by  $\mathbf{r}(\lambda)$ . Then  $\mathbf{r}: \Delta \rightarrow [0, 1]$  is continuous.*

In order to prove Theorem 6.2 we shall use the following general lemma.

**Lemma 6.3.** *Let  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  be a compact set indexed in  $\mathcal{I}$ . For each  $n \geq 1$  let  $\mathbf{A}_n = \{A_i^{(n)}: i \in \mathcal{I}\}$  be a compact subset of  $\mathbf{M}_d(\mathbb{C})$  indexed in  $\mathcal{I}$ , and suppose that the uniform difference  $\sup \left\{ \|A_i - A_i^{(n)}\|: i \in \mathcal{I} \right\}$  converges to zero as  $n \rightarrow \infty$ . If  $\mu_n \in \mathcal{M}_{\max}(\mathbf{A}_n)$  for each  $n \geq 1$ , and  $\mu \in \mathcal{M}_{\sigma}$  is a limit point of  $(\mu_n)$ , then  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ .*

*Proof.* We shall make use of the fact that the convergence hypothesis on  $\mathbf{A}_n$  implies that  $\lim_{n \rightarrow \infty} \varrho(\mathbf{A}_n) = \varrho(\mathbf{A})$ . For a proof of this statement see e.g. [34, 61]. By passing to an appropriate subsequence if necessary, we may assume that in fact  $\mu$  is the weak-\* limit of the sequence  $(\mu_n)$ .

Let us fix two integers  $k, m \geq 1$ . It follows from our hypotheses that the sequence of continuous functions from  $\Sigma_{\mathcal{I}}$  to  $\mathbb{R}$  defined by  $x \mapsto \max\{\log \|\mathcal{L}_{\mathbf{A}_n}(x, m)\|, -k\}$  converges uniformly as  $n \rightarrow \infty$  to the continuous function defined by  $x \mapsto \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\}$ , and hence by a uniform estimate we may establish the equality

$$\lim_{n \rightarrow \infty} \left| \frac{1}{m} \int (\max\{\log \|\mathcal{L}_{\mathbf{A}_n}(x, m)\|, -k\} - \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\}) d\mu_n \right| = 0.$$

On the other hand, since  $(\mu_n)$  converges weakly to  $\mu$ , we have

$$\lim_{n \rightarrow \infty} \left| \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\} d\mu_n - \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\} d\mu \right| = 0$$

and combining these two estimates we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}_n}(x, m)\|, -k\} d\mu_n = \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\} d\mu$$

for each  $m$  and  $k$ . Hence for each  $m, k \geq 1$ ,

$$\begin{aligned} \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\} d\mu &= \lim_{n \rightarrow \infty} \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}_n}(x, m)\|, -k\} d\mu_n \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{m} \int \log \|\mathcal{L}_{\mathbf{A}_n}(x, m)\| d\mu_n \\ &\geq \liminf_{n \rightarrow \infty} \log \varrho(\mathbf{A}_n) = \log \varrho(\mathbf{A}), \end{aligned}$$

since by hypothesis we have  $\mu_n \in \mathcal{M}_{\max}(\mathbf{A}_n)$ . Applying the monotone convergence theorem we find that for each  $m \geq 1$

$$\frac{1}{m} \int \log \|\mathcal{L}_{\mathbf{A}}(x, m)\| d\mu = \lim_{k \rightarrow \infty} \frac{1}{m} \int \max\{\log \|\mathcal{L}_{\mathbf{A}}(x, m)\|, -k\} d\mu \geq \log \varrho(\mathbf{A}),$$

and since  $m$  is arbitrary we conclude that  $\mu \in \mathcal{M}_{\max}(\mathbf{A})$ . The proof is complete.  $\square$

of Theorem 6.2. We shall show that if  $(\lambda_n)$  is a sequence of elements of  $\Delta$  converging to a limit  $\lambda \in \Delta$ , then  $\tau(\lambda_n)$  converges to  $\tau(\lambda)$ . Given such a sequence, for each  $n \geq 1$  let us choose a measure  $\mu_n \in \mathcal{M}_{\max}(\mathbf{A}_{\lambda_n})$ . Let  $(n_j)$  be a subsequence of the natural numbers such that  $\tau(\lambda_{n_j})$  converges as  $j \rightarrow \infty$  to a limit  $\gamma \in [0, 1]$ . Replacing  $(n_j)$  with a finer subsequence if necessary, and making use of the fact that  $\mathcal{M}_\sigma$  is compact and metrisable, we may assume that  $(\mu_{n_j})$  converges to a limit  $\mu \in \mathcal{M}_\sigma$ . By hypothesis we have  $\lim_{j \rightarrow \infty} \max\{\|A_k(\lambda) - A_k(\lambda_{n_j})\| : 1 \leq k \leq \ell\} = 0$ , so we may apply Lemma 6.3 to deduce that  $\mu \in \mathcal{M}_{\max}(\mathbf{A}_\lambda)$ . Since  $\chi_{[i]} : \Sigma_\ell \rightarrow \mathbb{R}$  is continuous and  $\tau(\lambda)$  is the  $i$ -ratio of  $\mathbf{A}_\lambda$ , by weak-\* convergence we have  $\tau(\lambda) = \mu([i]) = \lim_{j \rightarrow \infty} \int \chi_{[i]} d\mu_{n_j} = \lim_{j \rightarrow \infty} \tau(\lambda_{n_j}) = \gamma$ . We conclude that  $\tau(\lambda)$  is the only limit point of the bounded sequence  $(\tau(\lambda_n))$ , and therefore that sequence converges to  $\tau(\lambda)$ . The proof is complete.  $\square$

## 7. APPLICATIONS TO THE UNIQUENESS OF BARABANOV NORMS

In this section we shall say that  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  is *reducible* if there exists a nontrivial proper subspace of  $\mathbb{C}^d$  which is preserved by every element of  $\mathbf{A}$ ; otherwise we shall say that  $\mathbf{A}$  is *irreducible*. If  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  is a nonempty bounded set, we say that a norm  $\|\cdot\|$  on  $\mathbb{C}^d$  is a *Barabanov norm* if the property  $\varrho(\mathbf{A})\|v\| = \max_{A \in \mathbf{A}} \|Av\|$  is satisfied for all  $v \in \mathbb{C}^d$ . A fundamental result due to N. E. Barabanov demonstrates that if  $\mathbf{A} \subset \mathbf{M}_d(\mathbb{C})$  is compact and irreducible, then it admits at least one Barabanov norm [1]. The properties and applications of Barabanov norms were subsequently investigated in depth by F. Wirth and E. Plischke [48, 61, 62, 63], and algorithms for their computation have been suggested by V. S. Kozyakin [35, 36].

Since clearly any positive scalar multiple of a Barabanov norm is also a Barabanov norm, we shall say that  $\mathbf{A}$  has a ‘unique’ Barabanov norm to mean that all Barabanov norms for  $\mathbf{A}$  are directly proportional to one another. In [43], the present author investigated the question of when a finite irreducible set of matrices  $\mathbf{A}$  admits a unique Barabanov norm. In this section we show that a sufficient condition given in that article for a finite set  $\mathbf{A}$  to have a unique Barabanov norm can be described in terms of the associated Mather set  $Z_{\mathbf{A}}$ . This allows us to prove uniqueness of the Barabanov norm for a one-parameter family of pairs of matrices which was studied intensively in the articles [7, 25, 57].

Let  $\mathbf{A} = \{A_1, \dots, A_m\} \subset \mathbf{M}_d(\mathbb{C})$ . In [43], we said that  $\mathbf{A}$  has the *unbounded agreements property* if the following property holds: if  $j_1, j_2 : \mathbb{N} \rightarrow \{1, \dots, m\}$  are sequences such that  $\limsup_{n \rightarrow \infty} \|A_{j_1(n)} \cdots A_{j_1(1)}\| \varrho(\mathbf{A})^{-n} > 0$ , then for each  $k \geq 1$  there exist  $\ell_1, \ell_2 \geq 1$  such that  $j_1(\ell_1 + i) = j_2(\ell_2 + i)$  for all  $1 \leq i \leq k$ . This definition is motivated by the following result:

**Theorem 7.1** ([43]). *Let  $\mathbf{A} = \{A_1, \dots, A_m\} \subset \mathbf{M}_d(\mathbb{C})$ . Suppose that  $\mathbf{A}$  is irreducible and has the unbounded agreements property, and the set of matrices  $\wedge^2 \mathbf{A} := \{\wedge^2 A_1, \dots, \wedge^2 A_m\}$  satisfies  $\varrho(\wedge^2 \mathbf{A}) < \varrho(\mathbf{A})^2$ . Then  $\mathbf{A}$  has a unique Barabanov norm.*

The unbounded agreements property admits the following description in terms of the Mather set  $Z_{\mathbf{A}}$ :

**Proposition 7.2.** *Let  $\mathbf{A} = \{A_1, \dots, A_m\} \subset \mathbf{M}_d(\mathbb{C})$ . Then  $\mathbf{A}$  has the unbounded agreements property if and only if  $Z_{\mathbf{A}} \subseteq \Sigma_m$  contains a unique minimal set.*

*Proof.* Let us suppose that  $Z_{\mathbf{A}}$  contains a unique minimal set, and choose sequences  $x^{(1)} = (x_j^{(1)})$  and  $x^{(2)} = (x_j^{(2)}) \in \Sigma_m$  such that  $\limsup_{n \rightarrow \infty} \|\mathcal{L}_{\mathbf{A}}(x^{(i)}, n)\| \varrho(\mathbf{A})^{-n} > 0$  for  $i = 1, 2$ . Let us write  $Y_i := \overline{\{\sigma^\ell x^{(i)} : \ell \geq 0\}}$  for each  $i$ . Clearly,

$$\limsup_{n \rightarrow \infty} \sup_{y \in Y_i} \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(y, n)\| \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(x^{(i)}, n)\| = \log \varrho(\mathbf{A})$$

for  $i = 1, 2$ , and consequently

$$\lim_{n \rightarrow \infty} \sup_{y \in Y_i} \frac{1}{n} \log \|\mathcal{L}_{\mathbf{A}}(y, n)\| = \log \varrho(\mathbf{A})$$

since this limit exists by Lemma A.1, and its value can be at most  $\log \varrho(\mathbf{A})$  by the definition of  $\varrho(\mathbf{A})$ . Applying Theorem A.3 with  $X := Y_i$  and  $f_n(x) := \log \|\mathcal{L}_{\mathbf{A}}(x, n)\|$  we deduce that for each  $i = 1, 2$  there exists a  $\sigma$ -invariant Borel probability measure  $\mu_i$  on  $Y_i$  such that  $\inf_{m \geq 1} \frac{1}{m} \int \log \|\mathcal{L}_{\mathbf{A}}(x, m)\| d\mu_i(x) = \log \varrho(\mathbf{A})$ . In particular each  $\mu_i$  belongs to  $\mathcal{M}_{\max}(\mathbf{A})$ , and it follows that for each  $i$  the set  $\text{supp } \mu_i \subseteq Y_i$  is a subset of  $Z_{\mathbf{A}}$ . It follows from a classic theorem of Birkhoff ([31, p.130]) that for each  $i$  the compact  $\sigma$ -invariant set  $\text{supp } \mu_i$  contains a minimal set. Since by hypothesis there is exactly one minimal set contained in  $Z_{\mathbf{A}}$ , we conclude that  $\text{supp } \mu_1 \cap \text{supp } \mu_2$  is nonempty. Choose any point  $z \in \text{supp } \mu_1 \cap \text{supp } \mu_2 \subseteq Y_1 \cap Y_2$ , and let  $k \geq 1$  be any integer. By definition every element of  $Y_1$  belongs to the closure of  $\{\sigma^\ell x^{(1)} : \ell \geq 0\}$ , and therefore there exists  $\ell_1 \geq 1$  such that  $d(\sigma^{\ell_1} x^{(1)}, z) < 2^{-k}$ , implying that  $x_{\ell_1+j}^{(1)} = z_j$  for  $j = 1, \dots, k$ . Equally there must exist  $\ell_2 \geq 1$  such that  $d(\sigma^{\ell_2} x^{(2)}, z) < 2^{-k}$ , and it follows that  $x_{\ell_2+j}^{(2)} = z_j = x_{\ell_1+j}^{(1)}$  for  $j = 1, \dots, k$ . Since  $k$  is arbitrary we conclude that  $\mathbf{A}$  has the unbounded agreements property as required.

Let us now prove the converse direction. By the aforementioned theorem of Birkhoff,  $Z_{\mathbf{A}}$  contains at least one minimal set. Let us suppose that  $Z_{\mathbf{A}}$  contains two distinct minimal sets,  $Z_1$  and  $Z_2$ . If  $Z_1 \cap Z_2$  were nonempty it would be a  $\sigma$ -invariant closed proper subset of either  $Z_1$  or  $Z_2$ , contradicting minimality, and it follows that  $Z_1$  and  $Z_2$  must be pairwise disjoint. Choose any  $x^{(1)} \in Z_1$  and  $x^{(2)} \in Z_2$ . By Theorem 2.3 each  $x^{(i)}$  is strongly extremal and hence satisfies  $\limsup_{n \rightarrow \infty} \|\mathcal{L}_{\mathbf{A}}(x^{(i)}, n)\| \varrho(\mathbf{A})^{-n} > 0$ . Since  $Z_1$  and  $Z_2$  are compact and do not intersect, we may choose  $k > 0$  such that  $d(y, z) > 2^{-k}$  whenever  $y \in Z_1$  and  $z \in Z_2$ . In particular, since every point of the form  $\sigma^j x^{(i)}$  belongs to  $Z_i$ , we have  $d(\sigma^{\ell_1} x^{(1)}, \sigma^{\ell_2} x^{(2)}) > 2^{-k}$  for all  $\ell_1, \ell_2 \geq 1$ . It follows that for every  $\ell_1, \ell_2 \geq 1$  we must have  $x_{\ell_1+j}^{(1)} \neq x_{\ell_2+j}^{(2)}$  for some  $j$  in the range  $1, \dots, k$ , and as such  $\mathbf{A}$  does not have the unbounded agreements property. The proof is complete.  $\square$

We immediately deduce the following:

**Corollary 7.3.** *Let  $\mathbf{A} = \{A_1, \dots, A_m\} \subset \mathbf{M}_d(\mathbb{C})$  and suppose that  $\mathcal{M}_{\max}(\mathbf{A})$  contains exactly one measure. Then  $\mathbf{A}$  has the unbounded agreements property.*

*Proof.* If  $\mathbf{A}$  does not have the unbounded agreements property, then by Proposition 7.2 there exist two distinct minimal sets  $Y_1, Y_2 \subseteq Z_{\mathbf{A}}$ . By the Krylov-Bogolioubov theorem there exist  $\sigma$ -invariant Borel probability measures  $\mu_1$  and  $\mu_2$  supported in  $Y_1$  and  $Y_2$  respectively. Since two distinct minimal sets cannot intersect one another we have  $\text{supp } \mu_1 \cap \text{supp } \mu_2 = \emptyset$  and so in particular  $\mu_1 \neq \mu_2$ . However, by Theorem 2.3 we have  $\mu_1, \mu_2 \in \mathcal{M}_{\max}(\mathbf{A})$  since  $\mu_i(Z_{\mathbf{A}}) = 1$  for  $i = 1, 2$ , and since  $\mu_1 \neq \mu_2$  this contradicts our hypothesis.  $\square$

In [43] we proved uniqueness of the Barabanov norm for certain sets of matrices constructed specifically for that purpose. Using Proposition 7.2 we are able to give the following more natural family of examples:

**Theorem 7.4.** *For each  $\alpha > 0$  define  $\mathbf{A}_\alpha = \{A_1, \alpha A_2\}$  where*

$$A_1 := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A_2 := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

*Then for each  $\alpha \in (0, 1]$  the pair  $\mathbf{A}_\alpha$  has a unique Barabanov norm.*

*Proof.* By computing the eigenspaces of each matrix one may establish that each of the pairs  $\mathbf{A}_\alpha$  is irreducible, and hence in particular has at least one Barabanov norm. It was shown in [25, Theorem 2.3] that for each  $\alpha \in (0, 1]$  there exists a unique minimal set  $X_{\mathbf{r}(\alpha)} \subset \Sigma_2$  such that if  $x \in \Sigma_2$  is recurrent and strongly extremal for  $\mathbf{A}_\alpha$ , then  $x \in X_{\mathbf{r}(\alpha)}$ . Since  $Z_{\mathbf{A}_\alpha}$  is closed and  $\sigma$ -invariant, it contains at least one minimal set. If  $Y \subseteq Z_{\mathbf{A}_\alpha}$  is minimal and  $x \in Y$  then  $x$  is recurrent by minimality and strongly extremal by Theorem 2.3, and it follows that  $x \in X_{\mathbf{r}(\alpha)}$ . In particular  $Y \cap X_{\mathbf{r}(\alpha)} \neq \emptyset$  and therefore  $Y = X_{\mathbf{r}(\alpha)}$  by minimality. We conclude that  $Z_{\mathbf{A}_\alpha}$  contains a unique minimal set, and by Proposition 7.2 it follows that  $\mathbf{A}_\alpha$  has the unbounded agreements property.

To prove the proposition it remains only to show that the joint spectral radius of the set  $\wedge^2 \mathbf{A}_\alpha := \{\wedge^2 A_1, \wedge^2 (\alpha A_2)\}$  is strictly less than  $\varrho(\mathbf{A}_\alpha)^2$ . However, since the matrices  $A_1, A_2$  are two-dimensional,  $\wedge^2 \mathbf{A}_\alpha$  simply consists of the two one-dimensional matrices with entries respectively equal to  $1 = \det A_1$  and  $\alpha = \det(\alpha A_2)$ . It follows that  $\varrho(\wedge^2 \mathbf{A}_\alpha) = 1$ . On the other hand by [25, Lemma 3.4] we have  $\varrho(\mathbf{A}_\alpha) > 1$  for each  $\alpha \in (0, 1]$ . We conclude that the conditions of Theorem 7.1 are met, and for each  $\alpha \in (0, 1]$  the pair  $\mathbf{A}_\alpha$  admits a unique Barabanov norm.  $\square$

*Remark.* It was shown in [7, 25] that for certain choices of  $\alpha \in (0, 1]$  the pair  $\mathbf{A}_\alpha$  does not have the finiteness property. The above proposition therefore implies the existence of a pair of matrices which admits a unique Barabanov norm but does not have the finiteness property. This result was previously claimed without proof in the article [43].

#### APPENDIX A. SUBADDITIVE ERGODIC OPTIMISATION

Throughout this section, we let  $T: X \rightarrow X$  be a continuous transformation of a compact metric space. We denote by  $\mathcal{M}$  the set of all Borel probability measures on  $X$ . We equip this set with the weak-\* topology, which is the smallest topology such that  $\mu \mapsto \int g \, d\mu$  is a continuous map from  $\mathcal{M}$  to  $\mathbb{R}$  for every  $g \in C(X)$ . Under the weak-\* topology  $\mathcal{M}$  is a compact metrisable space (see e.g. [4, 45]). We let  $\mathcal{M}_T \subseteq \mathcal{M}$  denote the set of all measures which are invariant with respect to  $T$ , and we define  $\mathcal{E}_T \subseteq \mathcal{M}_T$  to be the set of all invariant measures with respect to which  $T$  is ergodic.

*Ergodic optimisation* is concerned with the following general problem: given a continuous function  $f: X \rightarrow \mathbb{R}$ , what can we say about the maximum ergodic average  $\beta(f) := \max_{\mu \in \mathcal{M}_T} \int f \, d\mu$ , the set of measures which attain this average, and the set of points  $x \in X$  for which  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \beta(f)$ ? An overview of this research area may be found in [28]; some articles in this area of particular note include [10, 12, 15, 64]. In this appendix we are concerned with the more general

problem of identifying points  $x \in X$  and measures which maximise the growth of *subadditive* ergodic averages, which we describe in detail below. The theorems proved in this section mainly consist of extensions of results already present in the literature, and as such are not of the first order of originality; on the other hand, these results are sufficiently far removed from their antecedents in the literature that we feel it would be unreasonable to ask the reader to accept them without proof. For this reason we have decided to present them as an appendix separate from the main body of the article.

Recall that a sequence  $(a_n)$  of elements of  $\mathbb{R} \cup \{-\infty\}$  is called *subadditive* if one has  $a_{n+m} \leq a_n + a_m$  for all  $n, m \geq 1$ . The following classical result due to Fekete [21] is used numerous times throughout this article. We include the proof in view of its brevity.

**Lemma A.1.** *Let  $(a_n)$  be a subadditive sequence of elements of  $\mathbb{R} \cup \{-\infty\}$ . Then  $\lim_{n \rightarrow \infty} a_n/n = \inf_{n \geq 1} a_n/n \in \mathbb{R} \cup \{-\infty\}$ .*

*Proof.* Given a real number  $\lambda > \inf_{n \geq 1} a_n/n$ , choose an integer  $m$  such that  $a_m \leq m\lambda$ . For each  $n \geq 1$  write  $n = q_n m + r_n$  where  $q_n := \lfloor n/m \rfloor$  and  $0 \leq r_n < m$ . We have  $a_n \leq q_n a_m + a_{r_n} \leq n\lambda + r_n a_1$  for all  $n \geq 1$ , and it follows that  $\limsup_{n \rightarrow \infty} a_n/n \leq \lambda$ . Since  $\lambda$  is arbitrary we conclude that  $\limsup_{n \rightarrow \infty} a_n/n \leq \inf_{n \geq 1} a_n/n$ , and the result follows.  $\square$

A sequence  $(f_n)$  of functions from  $X$  to  $\mathbb{R} \cup \{-\infty\}$  will be called *subadditive* if for each  $x \in X$  and  $n, m \geq 1$  the inequality  $f_{n+m}(x) \leq f_n(T^m x) + f_m(x)$  is satisfied. A starting point for our study of subadditive sequences of functions is the following classical result:

**Theorem A.2** ((Subadditive ergodic theorem)). *Let  $(X, \mathcal{F}, \mu)$  be a probability space equipped with an ergodic measure-preserving transformation  $T$ . Suppose that  $(f_n)$  is a sequence of measurable functions from  $X$  to  $\mathbb{R} \cup \{-\infty\}$  such that  $\max\{f_n, 0\}$  is integrable for each  $n \geq 1$ , and such that for each  $n, m \geq 1$  one has  $f_{n+m}(x) \leq f_n(T^m x) + f_m(x)$  for  $\mu$ -a.e.  $x$ . Then  $\lim_{n \rightarrow \infty} f_n(x)/n = \inf_{n \geq 1} (1/n) \int f_n d\mu$  for  $\mu$ -a.e.  $x$ .*

*Remark.* It is more usual to state this theorem under the additional hypotheses that each  $f_n$  is integrable, and  $\inf_{n \geq 1} (1/n) \int f_n d\mu > -\infty$ . Some proofs of Theorem A.2 in this form may be found in e.g. [19, 32]. To deduce the version of Theorem A.2 given above, it suffices to show that if the additional hypotheses do not hold then necessarily  $\limsup_{n \rightarrow \infty} f_n(x)/n = -\infty$  a.e., which may be achieved by taking upper estimates using the Birkhoff ergodic theorem.

Theorem A.2 indicates that given a subadditive sequence of upper semi-continuous functions  $(f_n)$  and an invariant measure  $\mu \in \mathcal{M}_T$ , the quantity  $\inf_{n \geq 1} \frac{1}{n} \int f_n d\mu$  may be understood as the ergodic average of  $(f_n)$  with respect to  $\mu$ . This motivates the following definition: given a transformation  $T: X \rightarrow X$  and a subadditive sequence of upper semi-continuous functions  $(f_n)$  from  $X$  to  $\mathbb{R} \cup \{-\infty\}$ , we define the *maximum ergodic average* of the sequence  $(f_n)$  to be the quantity

$$\beta[(f_n)] := \sup_{\mu \in \mathcal{M}_T} \inf_{n \geq 1} \frac{1}{n} \int f_n d\mu.$$

We shall say that  $\mu \in \mathcal{M}_T$  is a *maximising measure* for  $(f_n)$  if  $\inf_{n \geq 1} \frac{1}{n} \int f_n d\mu = \beta[(f_n)]$ , and denote the set of all maximising measures by  $\mathcal{M}_{\max}[(f_n)]$ . The key objective of this section is to prove the following characterisation of  $\beta[(f_n)]$ :

**Theorem A.3.** *Let  $(f_n)$  be a subadditive sequence of upper semi-continuous functions taking values in  $\mathbb{R} \cup \{-\infty\}$ . Then,*

$$\begin{aligned}\beta[(f_n)] &= \sup_{\mu \in \mathcal{E}_T} \inf_{n \geq 1} \frac{1}{n} \int f_n d\mu = \inf_{n \geq 1} \sup_{x \in X} \frac{1}{n} f_n(x) \\ &= \inf_{n \geq 1} \sup_{\mu \in \mathcal{M}_T} \frac{1}{n} \int f_n d\mu = \sup_{x \in X} \inf_{n \geq 1} \frac{1}{n} f_n(x).\end{aligned}$$

*In all but the last of these expressions, the infimum over all  $n \geq 1$  may be replaced with the limit as  $n \rightarrow \infty$  of the same quantity, without altering the value of the expression. Furthermore, every supremum arising in each of the above expressions is attained.*

Theorem A.3 may be seen as a subadditive analogue of [28, Proposition 2.1]. Theorem A.3 extends previous results in the subadditive context which deal with the case in which each  $f_n$  is continuous and takes values only in  $\mathbb{R}$ . Under these additional hypotheses the identity

$$(9) \quad \beta[(f_n)] = \sup_{\mu \in \mathcal{E}_T} \inf_{n \geq 1} \frac{1}{n} \int f_n d\mu = \inf_{n \geq 1} \sup_{x \in X} \frac{1}{n} f_n(x) = \sup_{x \in X} \limsup_{n \rightarrow \infty} \frac{1}{n} f_n(x)$$

was previously established by S. J. Schreiber [53]. The same relationships were later independently proved by R. Sturman and J. Stark under the same hypotheses [55]. Our removal of the condition that each  $f_n$  takes only real values has the significant advantage that we may treat linear cocycles of matrices or operators which may not be invertible or even nonzero at every point. We also remove the assumption that each  $f_n$  is lower semi-continuous. The attainment of the suprema in (9), which was not addressed in the work of Schreiber and Sturman-Stark, is also crucial in several of our applications. The identities

$$\beta[(f_n)] = \sup_{x \in X} \inf_{n \geq 1} \frac{1}{n} f_n(x) = \inf_{n \geq 1} \sup_{x \in X} \frac{1}{n} f_n(x)$$

and the attainment of the corresponding suprema are therefore original with this document.

In proving Theorem A.3 our approach essentially follows the method of Schreiber, as opposed to the somewhat different path taken by Sturman and Stark. This is motivated by the fact that this proof takes us naturally through a result - Lemma A.6 below - which has independent interest, and is applied separately from Theorem A.3 in the main document. However, our proof differs from those given by Schreiber and Sturman-Stark in that we do not actually make use of the subadditive ergodic theorem in proving Theorem A.3.

Before proving the identities which form the main part of Theorem A.3, we shall prove some basic properties of the set of maximising measures  $\mathcal{M}_{\max}[(f_n)]$ . We begin with the following simple lemma:

**Lemma A.4.** *If  $g: X \rightarrow \mathbb{R} \cup \{-\infty\}$  is upper semi-continuous, then the map from  $\mathcal{M}$  to  $\mathbb{R} \cup \{-\infty\}$  given by  $\mu \mapsto \int g d\mu$  is upper semi-continuous. If  $(f_n)$  is a subadditive sequence of upper semi-continuous functions from  $X$  to  $\mathbb{R} \cup \{-\infty\}$ , then the map from  $\mathcal{M} \rightarrow \mathbb{R} \cup \{-\infty\}$  given by  $\mu \mapsto \inf_{m \geq 1} \frac{1}{m} \int f_m d\mu$  is also upper semi-continuous.*

*Proof.* A function from a metrisable space to  $\mathbb{R} \cup \{-\infty\}$  is upper semi-continuous if and only if it is equal to the pointwise limit of a decreasing sequence of continuous

functions taking values in  $\mathbb{R}$  (see e.g. [9, ch. IX]). Given a function  $g$  as above, let  $(g_i)_{i=1}^\infty$  be such a decreasing sequence converging pointwise to  $g$ . For each  $i$  the map  $\mu \mapsto \int g_i d\mu$  is clearly real-valued, and is by definition weak-\* continuous. For each  $\mu \in \mathcal{M}$  the real-valued sequence  $(\int g_i d\mu)_{i=1}^\infty$  decreases to  $\int g d\mu$  by the monotone convergence theorem, and it follows that the map  $\mu \mapsto \int g d\mu$  is upper semi-continuous. If  $(f_n)$  is a subadditive sequence as above then each of the functions  $\mu \mapsto \frac{1}{m} \int f_m d\mu$  is upper semi-continuous by the preceding argument, and since the pointwise infimum of a family of upper semi-continuous functions is also upper semi-continuous the result follows.  $\square$

We immediately deduce:

**Proposition A.5.** *Let  $(f_n)$  be a subadditive sequence of upper semi-continuous functions taking values in  $\mathbb{R} \cup \{-\infty\}$ . Then  $\mathcal{M}_{\max}[(f_n)]$  is compact, convex and nonempty, and the extreme points of  $\mathcal{M}_{\max}[(f_n)]$  are precisely the ergodic elements of  $\mathcal{M}_{\max}[(f_n)]$ .*

*Proof.* It follows from the semicontinuity of the map  $\mu \mapsto \inf_{m \geq 1} \frac{1}{m} \int f_m d\mu$  that  $\mathcal{M}_{\max}(f_m)$  is compact and nonempty. By subadditivity we have  $\mu \in \mathcal{M}_{\max}[(f_m)]$  if and only if  $\lim_{n \rightarrow \infty} \frac{1}{n} \int f_n d\mu = \beta[(f_n)]$ , and it is clear from this that if  $m_1, m_2 \in \mathcal{M}_{\max}[(f_m)]$  then  $\lambda m_1 + (1 - \lambda)m_2 \in \mathcal{M}_{\max}[(f_m)]$  for every  $\lambda \in [0, 1]$ .

If  $\mu \in \mathcal{M}_{\max}[(f_n)]$  is ergodic then it is well-known that  $\mu$  is an extreme point of  $\mathcal{M}_T$  (see e.g. [59, p.152]), and hence is also an extreme point of  $\mathcal{M}_{\max}[(f_n)]$ . Conversely, if  $\mu$  is a non-ergodic maximising measure, let  $A \subseteq X$  be a Borel set such that  $T^{-1}A = A$  up to  $\mu$ -measure zero and  $0 < \mu(A) < 1$ . Define measures  $m_1, m_2 \in \mathcal{M}_T$  by  $m_1(Y) := \mu(A \cap Y)/\mu(A)$  and  $m_2(Y) := \mu(Y \setminus A)/\mu(X \setminus A)$ . An easy argument by contradiction shows that both  $m_1$  and  $m_2$  must belong to  $\mathcal{M}_{\max}[(f_m)]$ , and since  $\mu = \mu(A)m_1 + (1 - \mu(A))m_2$  it follows that  $\mu$  is not an extreme point of  $\mathcal{M}_{\max}[(f_n)]$ .  $\square$

The next two results together constitute the bulk of the proof of Theorem A.3, but each has additional usefulness in its own right. Lemma A.6 below is a modification of a result of Schreiber [53, Lemma 1], although some differences exist since parts of Schreiber's proof require the quantity  $\inf f_n$  to be finite for each  $n$ .

**Lemma A.6.** *Let  $(f_n)$  be a subadditive sequence of upper semi-continuous functions taking values in  $\mathbb{R} \cup \{-\infty\}$ , let  $x \in X$ , and suppose that  $\lim_{n \rightarrow \infty} (1/n)f_n(x) = c$ . Define a sequence of measures  $\mu_n \in \mathcal{M}$  by  $\mu_n := (1/n) \sum_{k=0}^{n-1} \delta_{T^k x}$ , where  $\delta_z$  denotes the Dirac measure concentrated at the point  $z \in X$ , and suppose that  $\mu \in \mathcal{M}$  is a weak-\* accumulation point of  $(\mu_n)$ . Then  $\mu \in \mathcal{M}_T$  and  $\inf_{m \geq 1} \frac{1}{m} \int f_m d\mu \geq c$ .*

*Proof.* The proof that every accumulation point of  $(\mu_n)$  must lie in  $\mathcal{M}_T$  is entirely standard, see for example [59, p.151]. If  $c = -\infty$  then this suffices to complete the proof, so for the remainder of the proof we assume that  $c$  is real. To prove that  $\inf_{m \geq 1} \frac{1}{m} \int f_m d\mu \geq c$  for every accumulation point  $\mu$ , it suffices to prove that  $\liminf_{n \rightarrow \infty} \frac{1}{n} \int f_n d\mu_n \geq c$  for every  $n \geq 1$  and apply the first clause of Lemma A.4.

We begin with the following claim: if  $k, \ell \geq 1$  then  $\lim_{n \rightarrow \infty} \frac{1}{n} f_k(T^{n-\ell}x) = 0$ . To see this, note that for large enough  $n$  subadditivity yields  $f_{n+k-\ell}(x) \leq f_{n-\ell}(x) + f_k(T^{n-\ell}x) \leq f_{n-\ell}(x) + \sup f_k$ , and since clearly  $\lim_{n \rightarrow \infty} \frac{1}{n} f_{n+k-\ell}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} (f_{n-\ell}(x) + \sup f_k) = c$  we obtain the desired result.

Let us now fix  $m \geq 1$ , and show that  $\liminf_{n \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^{n-1} f_m(T^i x) \geq c$ . Since

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{nm} \left( \sum_{i=0}^{n-1} f_m(T^i x) - \sum_{i=0}^{n-m+1} f_m(T^i x) \right) &= \lim_{n \rightarrow \infty} \frac{1}{nm} \sum_{i=n-m+2}^{n-1} f_m(T^i x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{nm} \sum_{i=1}^{m-2} f_m(T^{n-i} x) = 0 \end{aligned}$$

by our previous claim, it suffices to show that  $\liminf_{n \rightarrow \infty} \frac{1}{nm} \sum_{i=0}^{n-m+1} f_m(T^i x) \geq c$ . For each  $i$  in the range  $0 \leq i < m$ , then, let us choose integers  $q_i, r_i$  such that  $n = i + q_i m + r_i$  with  $q_i \geq 0$  and  $0 \leq r_i < m$ . We have

$$\sum_{i=0}^{m-1} \sum_{j=0}^{q_i-1} f_m(T^{i+jm} x) = \sum_{i=0}^{n-m+1} f_m(T^i x),$$

and so using subadditivity we may estimate

$$\begin{aligned} mf_n(x) &\leq \sum_{i=0}^{m-1} \left( f_i(x) + \sum_{j=0}^{q_i-1} f_m(T^{i+jm} x) + f_{r_i}(T^{i+q_i m} x) \right) \\ &= \sum_{i=0}^{m-1} f_i(x) + \sum_{i=0}^{n-m+1} f_m(T^i x) + \sum_{i=0}^{m-1} f_{r_i}(T^{n-r_i} x). \end{aligned}$$

Dividing both sides by  $nm$ , taking the limit inferior as  $n \rightarrow \infty$  and applying our previous claim finishes the proof.  $\square$

**Proposition A.7.** *Let  $(f_n)$  be a subadditive sequence of upper semi-continuous functions taking values in  $\mathbb{R} \cup \{-\infty\}$ , and define  $\tilde{\beta}[(f_n)] := \lim_{n \rightarrow \infty} \sup_{x \in X} \frac{1}{n} f_n(x)$ . Then there exists  $z \in X$  such that  $\inf_{m \geq 1} \frac{1}{m} f_m(z) \geq \tilde{\beta}[(f_n)]$ .*

*Proof.* To simplify some expressions in this proof we shall denote the value  $\tilde{\beta}[(f_n)]$  simply by  $\tilde{\beta}$ . If  $\tilde{\beta} = -\infty$  then it suffices to choose any  $z \in X$ , so we shall assume that this value is finite. To prove the Proposition we shall assume that no point  $z \in X$  exists with the required property, and show that this leads to a contradiction.

Making this hypothesis, it follows that for every  $x \in X$  there exists an integer  $n(x) \geq 1$  and real number  $\varepsilon(x) > 0$  such that  $f_{n(x)}(x) < n(x)(\tilde{\beta} - \varepsilon(x))$ . Since each  $f_k$  is upper semi-continuous, it follows that for each  $x \in X$  there exists an open subset  $U_x$  of  $X$  containing  $x$  such that  $f_{n(x)}(y) < n(x)(\tilde{\beta} - \varepsilon(x))$  for all  $y \in U_x$ . By compactness we may choose finitely many points  $x_1, \dots, x_k \in X$  such that  $\bigcup_{i=1}^k U_{x_i} = X$ . It follows that we may choose a real number  $\varepsilon > 0$  and a function  $r_1: X \rightarrow \mathbb{N}$  taking values in the set  $\{n(x_1), n(x_2), \dots, n(x_k)\}$  such that  $f_{r_1(x)}(x) < r_1(x)(\tilde{\beta} - \varepsilon)$  for every  $x \in X$ . Let  $m, M \geq 1$  be natural numbers such that  $m \leq r_1(x) < M$  for every  $x$ .

Using the function  $r_1$ , we shall now define a sequence of functions  $r_k: X \rightarrow \mathbb{N}$  for all  $k \geq 2$ . The function  $r_k$  having been defined, let us define  $r_{k+1}$  by setting  $r_{k+1}(x) := r_k(x) + r_1(T^{r_k(x)} x)$  for all  $x \in X$ . An elementary induction shows that  $km \leq r_k(x) \leq k(M-1)$  for every  $x \in X$  and  $k \geq 1$ . We claim that additionally  $f_{r_k(x)}(x) < r_k(x)(\tilde{\beta} - \varepsilon)$  for all  $x \in X$  and  $k \geq 1$ . In the case  $k = 1$  this has already been established. Given that this relation holds for all  $x \in X$  for some

fixed  $k \geq 1$ , note that for each  $x \in X$  we have

$$\begin{aligned} f_{r_{k+1}(x)}(x) &\leq f_{r_1(T^{r_k(x)}x)}\left(T^{r_k(x)}x\right) + f_{r_k(x)}(x) \\ &\leq r_1\left(T^{r_k(x)}x\right)\left(\tilde{\beta} - \varepsilon\right) + r_k(x)\left(\tilde{\beta} - \varepsilon\right) = r_{k+1}(x)\left(\tilde{\beta} - \varepsilon\right), \end{aligned}$$

which establishes the required inequality for all  $x \in X$  in the case  $k+1$ . The result follows for all  $x \in X$  and  $k \geq 1$  by induction on  $k$ .

Now, for each  $k \geq 1$  define  $\tilde{\beta}_k := \sup_{i \geq k} \sup_{x \in X} \frac{1}{i} f_i(x)$  and note that the sequence  $(\tilde{\beta}_k)$  decreases to the limit  $\tilde{\beta}$  as  $k \rightarrow \infty$ . Define also  $\ell_k(x) := kM - r_k(x)$  for every  $k \geq 1$  and  $x \in X$ , and note that  $k \leq \ell_k(x) \leq k(M-m)$  as a consequence of the bounds on  $r_k(x)$ . For all  $x \in X$  and  $k \geq 1$  we have the inequality

$$\begin{aligned} f_{kM}(x) &\leq f_{r_k(x)}(x) + f_{\ell_k(x)}\left(T^{r_k(x)}x\right) \leq r_k(x)\left(\tilde{\beta} - \varepsilon\right) + \ell_k(x)\tilde{\beta}_k \\ &= kM\tilde{\beta} - r_k(x)\varepsilon + \ell_k(x)(\tilde{\beta}_k - \tilde{\beta}) \leq kM\tilde{\beta} - km\varepsilon + k(M-m)(\tilde{\beta}_k - \tilde{\beta}). \end{aligned}$$

Taking the supremum over  $x \in X$ , dividing both sides by  $kM$  and taking the limit as  $k \rightarrow \infty$  we find that  $\sup_{x \in X} \frac{1}{kM} f_{kM}(x) < \tilde{\beta}$  for all large enough  $k$ , which is a contradiction. The proof is complete.  $\square$

*Remark.* Proposition A.7 generalises a lemma of Y. Peres; previous proofs of that lemma relied on the maximal ergodic theorem [46, 50]. Note that the point  $z$  guaranteed by Proposition A.7 may be unique: if  $T: X \rightarrow X$  is minimal, then for the sequence  $(f_n)$  defined by  $f_n(x) := d(T^n x, z_0) - d(x, z_0)$ , the only point  $z \in X$  such that  $f_m(z) \geq m\tilde{\beta}[(f_n)] = 0$  for all  $m \geq 1$  is  $z_0$ .

By modifying the standard proof of the existence of minimal subsets of dynamical systems, we can obtain the following interesting extension of Proposition A.7:

**Corollary A.8.** *Let  $(f_n)$  be as in Proposition A.7. Then without loss of generality the point  $z \in X$  described by Proposition A.7 may be taken to be recurrent.*

*Proof.* Let us denote by  $\mathfrak{H}$  the set of all nonempty compact sets  $Z \subseteq X$  such that both  $TZ \subseteq Z$  and  $\inf_{n \geq 1} \sup_{x \in Z} \frac{1}{n} f_n(x) = \tilde{\beta}[(f_n)]$ , and equip  $Z$  with the partial order given by set inclusion. We shall show using Zorn's lemma that  $\mathfrak{H}$  has a minimal element with respect to this ordering.

To this end, let us suppose that  $\mathfrak{O}$  is a subset of  $\mathfrak{H}$  in which set inclusion is a total order, and define  $Y := \bigcap_{Z \in \mathfrak{O}} Z$ . Clearly  $Y$  is compact and  $T$ -invariant, and it is nonempty since otherwise  $\bigcup_{Z \in \mathfrak{O}} (X \setminus Z)$  would be an open cover of  $X$  without a finite subcover. For each  $m \geq 1$  the set  $\{x \in X: f_m(x) \geq m\tilde{\beta}[(f_n)]\}$  is compact and intersects every  $Z \in \mathfrak{O}$ , so in particular each such set must intersect  $Y$ ; it follows from this that  $\inf_{m \geq 1} \sup_{x \in Y} \frac{1}{m} f_m(x) = \tilde{\beta}[(f_n)]$ . We conclude that  $Y \in \mathfrak{H}$  and  $Y \subseteq Z$  for every  $Z \in \mathfrak{O}$ , and it follows by Zorn's lemma that  $\mathfrak{H}$  must have a minimal element with respect to inclusion.

Let  $Z \in \mathfrak{H}$  be such a minimal element. By applying Proposition A.7 to  $Z$ , there exists  $z \in Z$  such that  $\inf_{m \geq 1} \frac{1}{m} f_m(z) \geq \tilde{\beta}[(f_n)]$ . We claim that  $z$  is recurrent. Let us define  $\tilde{Z} := \overline{\{T^k z: k \geq 1\}}$ . Clearly  $\tilde{Z}$  is closed and  $T$ -invariant. For each  $m \geq 1$  we have  $(m+1)\tilde{\beta}[(f_n)] \leq f_{m+1}(z) \leq f_1(z) + f_m(Tz)$ , and hence

$$\beta[(f_n)] \geq \inf_{m \geq 1} \sup_{x \in \tilde{Z}} \frac{1}{m} f_m(x) = \lim_{m \rightarrow \infty} \sup_{x \in \tilde{Z}} \frac{1}{m} f_m(x) \geq \liminf_{m \rightarrow \infty} \frac{1}{m} f_m(Tz) \geq \tilde{\beta}[(f_n)],$$

from which it follows that  $\tilde{Z} \in \mathfrak{H}$ . Since by assumption  $Z$  is a minimal element of  $\mathfrak{H}$  with respect to inclusion, and  $\tilde{Z} \subseteq Z$ , we conclude that necessarily  $\tilde{Z} = Z$ , and this implies that  $z$  is recurrent as claimed.  $\square$

of Theorem A.3. Combining respectively Proposition A.7, Lemma A.6 and Proposition A.5 we immediately obtain

(10)

$$\inf_{n \geq 1} \sup_{x \in X} \frac{1}{n} f_n(x) \leq \sup_{x \in X} \inf_{n \geq 1} \frac{1}{n} f_n(x) \leq \sup_{\mu \in \mathcal{M}_T} \inf_{n \geq 1} \frac{1}{n} \int f_n d\mu = \sup_{\mu \in \mathcal{E}_T} \inf_{n \geq 1} \frac{1}{n} \int f_n d\mu,$$

and the suprema in all of these expressions are attained. Now, if  $n \geq 1$  and  $\nu \in \mathcal{M}_T$ , the inequality  $\frac{1}{n} \int f_n d\nu \leq \sup_{\mu \in \mathcal{M}_T} \frac{1}{n} \int f_n d\mu \leq \sup_{x \in X} \frac{1}{n} f_n$  is obvious. Taking first the infimum over  $n \geq 1$  and then the supremum over  $\nu$  we obtain

$$(11) \quad \sup_{\nu \in \mathcal{M}_T} \inf_{n \geq 1} \frac{1}{n} \int f_n d\nu \leq \inf_{n \geq 1} \sup_{\mu \in \mathcal{M}_T} \frac{1}{n} \int f_n d\mu \leq \inf_{n \geq 1} \sup_{x \in X} \frac{1}{n} f_n(x),$$

and it follows from Lemma A.4 and the compactness of  $\mathcal{M}_T$  that the supremum in the middle expression is also attained for every  $n$ . Combining (10) and (11) proves all of the desired identities. Lastly, we note that with the single exception of the quantity  $\inf_{n \geq 1} \frac{1}{n} f_n(x)$ , every infimum arising in these expressions can be expressed in the form  $\inf_{n \geq 1} \frac{1}{n} a_n$  where  $(a_n)$  is a subadditive sequence, and hence by Lemma A.1 is also a limit. The proof is complete.  $\square$

In ergodic optimisation, a continuous function  $f: X \rightarrow \mathbb{R}$  is said to satisfy the *subordination principle*, introduced in [10], if the set  $\mathcal{M}_{\max}(f) := \{\mu \in \mathcal{M}_T: \int f d\mu = \beta(f)\}$  satisfies the following property: if  $\mu \in \mathcal{M}_{\max}(f)$ ,  $\nu \in \mathcal{M}_T$  and  $\text{supp } \nu \subseteq \text{supp } \mu$ , then  $\nu \in \mathcal{M}_{\max}(f)$ . This is equivalent to the existence of a compact  $T$ -invariant set  $Y \subseteq X$  such that if  $\mu \in \mathcal{M}_T$  and  $\text{supp } \mu \subseteq Y$ , then  $\mu \in \mathcal{M}_{\max}(f)$ . (For a proof of this result see [42]). The following result, which generalises an argument given in [44], describes an analogous phenomenon in the subadditive context.

**Lemma A.9.** *Let  $(f_n)$  be a subadditive sequence of upper semi-continuous functions taking values in  $\mathbb{R} \cup \{-\infty\}$ , and suppose that there exists  $\lambda \in \mathbb{R}$  such that  $\sup\{f_n(x): x \in X\} = n\lambda$  for infinitely many  $n \geq 1$ . Then  $\lambda = \beta[(f_n)]$ , and the set*

$$Y := \bigcap_{n=1}^{\infty} \{x \in X: f_n(x) = n\lambda\}$$

*is compact, nonempty, and satisfies  $TY \subseteq Y$ . Moreover, for each  $\mu \in \mathcal{M}_T$  we have  $\mu \in \mathcal{M}_{\max}[(f_n)]$  if and only if  $\text{supp } \mu \subseteq Y$ .*

*Proof.* For each  $n \geq 1$  define  $Y_n := \{x \in X: f_n(x) = n\lambda\}$ . Using semicontinuity and the fact that  $\sup f_n = n\lambda$  for each  $n$  it follows that each  $Y_n$  is nonempty and closed, hence compact. We claim that  $Y_{n+1} \subseteq Y_n$  and  $Y_{n+1} \subseteq T^{-1}Y_n$  for each  $n$ . To see the former, note that if  $x \in Y_{n+1}$  then  $(n+1)\lambda = f_{n+1}(x) \leq f_n(x) + f_1(T^n x) \leq f_n(x) + \lambda \leq (n+1)\lambda$  and therefore  $f_n(x) = n\lambda$  which implies that  $x \in Y_n$ . To see the latter we similarly observe that if  $x \in Y_{n+1}$ , then since  $(n+1)\lambda = f_{n+1}(x) \leq f_1(x) + f_n(Tx) \leq \lambda + f_n(Tx) \leq (n+1)\lambda$  we have  $f_n(Tx) = n\lambda$  and therefore  $Tx \in Y_n$ . It follows that the intersection  $Y := \bigcap_{n=1}^{\infty} Y_n$  is compact and nonempty as claimed, and  $TY \subseteq Y$ . We have  $\lambda = \beta[(f_n)]$  as a direct consequence of Theorem A.3. It follows immediately that if  $\mu \in \mathcal{M}_T$ , then  $\mu \in \mathcal{M}_{\max}[(f_n)]$  if and only if  $\int f_n d\mu = n\lambda$  for all  $n \geq 1$ , if and only if  $\mu(Y_n) = 1$  for all  $n \geq 1$ , if and only if  $\text{supp } \mu \subseteq Y$ .  $\square$

## REFERENCES

1. N. E. Barabanov, *On the Lyapunov exponent of discrete inclusions. I*, Automat. Remote Control **49** (1988), no. 2, 152–157. MR MR940263 (89e:34025)
2. G. P. Barker, L. T. Conner, Jr., and D. P. Stanford, *Complete controllability and contractibility in multimodal systems*, Linear Algebra Appl. **110** (1988), 55–74. MR 964208 (90a:93013)
3. Marc A. Berger and Yang Wang, *Bounded semigroups of matrices*, Linear Algebra Appl. **166** (1992), 21–27. MR MR1152485 (92m:15012)
4. Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968. MR 0233396 (38 #1718)
5. Vincent D. Blondel, Julien Cassaigne, and Raphaël M. Jungers, *On the number of  $\alpha$ -power-free binary words for  $2 < \alpha \leq 7/3$* , Theoret. Comput. Sci. **410** (2009), no. 30-32, 2823–2833. MR 2543336 (2010h:68141)
6. Vincent D. Blondel and Yurii Nesterov, *Computationally efficient approximations of the joint spectral radius*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 1, 256–272 (electronic). MR MR2176820 (2006k:15027)
7. Vincent D. Blondel, Jacques Theys, and Alexander A. Vladimirov, *An elementary counterexample to the finiteness conjecture*, SIAM J. Matrix Anal. Appl. **24** (2003), no. 4, 963–970 (electronic). MR MR2003315 (2004g:15010)
8. Jairo Bochi, *Inequalities for numerical invariants of sets of matrices*, Linear Algebra Appl. **368** (2003), 71–81. MR MR1983195 (2004f:15034)
9. Nicolas Bourbaki, *General topology. Chapters 5–10*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1998, Translated from the French, Reprint of the 1989 English translation. MR MR1726872 (2000h:54001b)
10. Thierry Bousch, *Le poisson n'a pas d'arêtes*, Ann. Inst. H. Poincaré Probab. Statist. **36** (2000), no. 4, 489–508. MR MR1785392 (2001i:37005)
11. ———, *La condition de Walters*, Ann. Sci. École Norm. Sup. (4) **34** (2001), no. 2, 287–311. MR 1841880 (2002h:37051)
12. Thierry Bousch and Oliver Jenkinson, *Cohomology classes of dynamically non-negative  $C^k$  functions*, Invent. Math. **148** (2002), no. 1, 207–217. MR 1892849 (2003c:37048)
13. Thierry Bousch and Jean Mairesse, *Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture*, J. Amer. Math. Soc. **15** (2002), no. 1, 77–111 (electronic). MR MR1862798 (2002j:49008)
14. Antonio Ciccone, Nicola Guglielmi, Stefano Serra-Capizzano, and Marino Zennaro, *Finiteness property of pairs of  $2 \times 2$  sign-matrices via real extremal polytope norms*, Linear Algebra Appl. **432** (2010), no. 2–3, 796–816. MR 2577718 (2011a:15047)
15. G. Contreras, A. O. Lopes, and Ph. Thieullen, *Lyapunov minimizing measures for expanding maps of the circle*, Ergodic Theory Dynam. Systems **21** (2001), no. 5, 1379–1409. MR MR1855838 (2002i:37038)
16. Xiongping Dai, Yu Huang, and Mingqing Xiao, *Almost sure stability of discrete-time switched linear systems: a topological point of view*, SIAM J. Control Optim. **47** (2008), no. 4, 2137–2156. MR 2421343 (2009i:93134)
17. ———, *Almost sure stability of discrete-time switched linear systems*, Proceedings of the 8th IEEE Conference on Control and Automation (Xiamen, China), June 2010, pp. 2098–2103.
18. Ingrid Daubechies and Jeffrey C. Lagarias, *Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal. **23** (1992), no. 4, 1031–1079. MR MR1166574 (93g:39001)
19. Yves Derriennic, *Sur le théorème ergodique sous-additif*, C. R. Acad. Sci. Paris Sér. A-B **281** (1975), no. 22, Aii, A985–A988. MR 0396903 (53 #763)
20. Lior Fainshil, Michael Margaliot, and Pavel Chigansky, *On the stability of positive linear switched systems under arbitrary switching laws*, IEEE Trans. Automat. Control **54** (2009), no. 4, 897–899. MR 2514831 (2010e:93048)
21. M. Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Math. Z. **17** (1923), no. 1, 228–249. MR 1544613
22. N. Guglielmi, F. Wirth, and M. Zennaro, *Complex polytope extremality results for families of matrices*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 3, 721–743 (electronic). MR MR2208331 (2007b:93106)

23. N. Guglielmi and M. Zennaro, *On the zero-stability of variable stepsize multistep methods: the spectral radius approach*, Numer. Math. **88** (2001), no. 3, 445–458. MR MR1835466 (2002d:65066)
24. Leonid Gurvits, *Stability of discrete linear inclusion*, Linear Algebra Appl. **231** (1995), 47–85. MR MR1361100 (96i:93056)
25. Kevin G. Hare, Ian David Morris, Nikita Sidorov, and Jacques Theys, *An explicit counterexample to the Lagarias-Wang finiteness conjecture*, Adv. Math. **226** (2011), 4667–4701.
26. Christopher Heil and Gilbert Strang, *Continuity of the joint spectral radius: application to wavelets*, Linear algebra for signal processing (Minneapolis, MN, 1992), IMA Vol. Math. Appl., vol. 69, Springer, New York, 1995, pp. 51–61.
27. Karl Heinrich Hofmann and Paul S. Mostert, *Elements of compact semigroups*, Charles E. Merr ll Books, Inc., Columbus, Ohio, 1966. MR MR0209387 (35 \#285)
28. Oliver Jenkinson, *Ergodic optimization*, Discrete Contin. Dyn. Syst. **15** (2006), no. 1, 197–224. MR MR2191393
29. Raphaël Jungers, *The joint spectral radius*, Lecture Notes in Control and Information Sciences, vol. 385, Springer-Verlag, Berlin, 2009, Theory and applications. MR 2507938
30. Raphaël M. Jungers and Vincent D. Blondel, *On the finiteness property for rational matrices*, Linear Algebra Appl. **428** (2008), no. 10, 2283–2295. MR MR2405245 (2009e:15029)
31. Anatole Katok and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995, With a supplementary chapter by Katok and Leonardo Mendoza. MR 1326374 (96c:58055)
32. Yitzhak Katznelson and Benjamin Weiss, *A simple proof of some ergodic theorems*, Israel J. Math. **42** (1982), no. 4, 291–296. MR 682312 (84i:28020)
33. Victor S. Kozyakin, *A dynamical systems construction of a counterexample to the finiteness conjecture*, Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005 (Seville, Spain), December 2005, pp. 2338–2343.
34. ———, *An explicit Lipschitz constant for the joint spectral radius*, Linear Algebra Appl. **433** (2010), no. 1, 12–18. MR 2645060
35. ———, *Iterative building of Barabanov norms and computation of the joint spectral radius for matrix sets*, Discrete Contin. Dyn. Syst. Ser. B **14** (2010), no. 1, 143–158. MR 2644257
36. ———, *Max-relaxation iteration procedure for building of Barabanov norms: convergence and examples*, arXiv preprint 1002.3251v2, 2010.
37. Jeffrey C. Lagarias and Yang Wang, *The finiteness conjecture for the generalized spectral radius of a set of matrices*, Linear Algebra Appl. **214** (1995), 17–42. MR MR1311628 (95k:15038)
38. Yuan-Chuan Li and Mau-Hsiang Shih, *The normed finiteness property of compact contraction operators*, Linear Algebra Appl. **428** (2008), no. 10, 2319–2323. MR 2405248 (2009k:47031)
39. Ricardo Mañé, *On the minimizing measures of Lagrangian dynamical systems*, Nonlinearity **5** (1992), no. 3, 623–638. MR 1166538 (93h:58059)
40. John N. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems*, Math. Z. **207** (1991), no. 2, 169–207. MR 1109661 (92m:58048)
41. Bruce E. Moision, Alon Orlitsky, and Paul H. Siegel, *On codes that avoid specified differences*, IEEE Trans. Inform. Theory **47** (2001), no. 1, 433–442. MR MR1820392 (2001k:94084)
42. Ian D. Morris, *A sufficient condition for the subordination principle in ergodic optimization*, Bull. Lond. Math. Soc. **39** (2007), no. 2, 214–220. MR MR2323451 (2008c:37052)
43. ———, *Criteria for the stability of the finiteness property and for the uniqueness of Barabanov norms*, Linear Algebra Appl. **433** (2010), no. 7, 1301 – 1311.
44. ———, *A rapidly-converging lower bound for the joint spectral radius via multiplicative ergodic theory*, Adv. Math. **225** (2010), 3425–3445.
45. K. R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967. MR 0226684 (37 \#2271)
46. Yuval Peres, *A combinatorial application of the maximal ergodic theorem*, Bull. London Math. Soc. **20** (1988), no. 3, 248–252. MR MR931186 (89e:28033)
47. Robert R. Phelps, *Lectures on Choquet’s theorem*, second ed., Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, Berlin, 2001. MR 1835574 (2002k:46001)
48. Elmar Plischke and Fabian Wirth, *Duality results for the joint spectral radius and transient behavior*, Linear Algebra Appl. **428** (2008), no. 10, 2368–2384. MR 2408033 (2009g:15033)

49. Vladimir Y. Protasov, Raphaël M. Jungers, and Vincent D. Blondel, *Joint spectral characteristics of matrices: a conic programming approach*, SIAM J. Matrix Anal. Appl. **31** (2009/10), no. 4, 2146–2162. MR 2678961
50. David Ralston, *Heaviness—an extension of a lemma of Y. Peres*, Houston J. Math. **35** (2009), no. 4, 1131–1141. MR MR2577147
51. Gian-Carlo Rota and Gilbert , *A note on the joint spectral radius*, Nederl. Akad. Wetensch. Proc. Ser. A 63 = Indag. Math. **22** (1960), 379–381. MR MR0147922 (26 \#5434)
52. Gian-Carlo Rota, *Gian-Carlo Rota on analysis and probability*, Contemporary Mathematicians, Birkhäuser Boston Inc., Boston, MA, 2003, Selected papers and commentaries, Edited by Jean Dhombres, Joseph P. S. Kung and Norton Starr. MR 1944526 (2004c:01047)
53. Sebastian J. Schreiber, *On growth rates of subadditive functions for semiflows*, J. Differential Equations **148** (1998), no. 2, 334–350. MR MR1643183 (2000a:37004)
54. Robert Shorten, Fabian Wirth, Oliver Mason, Kai Wulff, and Christopher King, *Stability criteria for switched and hybrid systems*, SIAM Rev. **49** (2007), no. 4, 545–592. MR 2375524 (2009e:93146)
55. R. Sturman and J. Stark, *Semi-uniform ergodic theorems and applications to forced systems*, Nonlinearity **13** (2000), no. 1, 113–143. MR MR1734626 (2000m:37041)
56. Zhendong Sun and S. S. Ge, *Analysis and synthesis of switched linear control systems*, Automatica J. IFAC **41** (2005), no. 2, 181–195. MR 2157653 (2006c:93056)
57. Jacques Theys, *Joint Spectral Radius: theory and approximations*, PhD thesis, Université Catholique de Louvain, 2005.
58. John N. Tsitsiklis and Vincent D. Blondel, *The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate*, Math. Control Signals Systems **10** (1997), no. 1, 31–40. MR MR1462278 (99h:65238a)
59. Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982. MR MR648108 (84e:28017)
60. Fabian Wirth, *Dynamics of time-varying discrete-time linear systems: spectral theory and the projected system*, SIAM J. Control Optim. **36** (1998), no. 2, 447–487. MR 1616502 (99i:93022)
61. ———, *The generalized spectral radius and extremal norms*, Linear Algebra Appl. **342** (2002), 17–40. MR MR1873424 (2003g:15025)
62. ———, *The generalized spectral radius is strictly increasing*, Linear Algebra Appl. **395** (2005), 141–153. MR MR2112880 (2005m:15027)
63. ———, *On the structure of the set of extremal norms of a linear inclusion*, Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005 (Seville, Spain), December 2005, pp. 3019–3024.
64. Guocheng Yuan and Brian R. Hunt, *Optimal orbits of hyperbolic systems*, Nonlinearity **12** (1999), no. 4, 1207–1224. MR MR1709845 (2000g:37029)